Foundations Research in Information Retrieval Inspired by Quantum Theory

Sachi Arafat

Department of Computer Science
Faculty of Information and Mathematical Sciences
University of Glasgow

UNIVERSITY
of
GLASGOW

Submitted for the Degree of Doctor of Philosophy
at the University of Glasgow
December 2007

© Sachi Arafat, 2007

Abstract

In the information age information is useless unless it can be found and used, search engines in our time thereby form a crucial component of research. For something so crucial, information retrieval (IR), the formal discipline investigating search, can be a confusing area of study. There is an underlying difficulty, with the very definition of information retrieval, and weaknesses in its operational method, which prevent it being called a science. The work in this thesis aims to create a formal definition for search, scientific methods for evaluation and comparison of different search strategies, and methods for dealing with the uncertainty associated with user interactions; so that one has the necessary formal foundation to be able to perceive IR as *"search science"*.

The key problems restricting a science of search pertain to the ambiguity in the current way in which search scenarios and concepts are specified. This especially affects evaluation of search systems since according to the traditional retrieval approach, evaluations are not repeatable, and thus not collectively verifiable. This is mainly due to the dependence on the method of user studies currently dominating evaluation methodology. This *evaluation problem* is related to the problem of not being able to formally define the users in user studies. The problem of defining users relates in turn to one of the main retrieval-specific motivations of the thesis, which can be understood by noticing that uncertainties associated with the interpretation of user interactions are collectively inscribed in a *relevance* concept, the representation and use of which defines the overall *character* of a retrieval model. Current research is limited in its understanding of how to best model relevance, a key factor restricting extensive formalization of the IR discipline as a whole. Thus, the problems of defining search systems and search scenarios are the principle issues preventing formal comparisons of systems and scenarios, in turn limiting the strength of experimental evaluation. Alternative models of search are proposed that remove the need for ambiguous relevance concepts and instead by arguing for use of simulation as a normative evaluation strategy for retrieval, some new concepts are introduced that can be employed in judging *effectiveness* of search systems. Included are techniques for simulating search, techniques for formal user modelling and techniques for generating measures of effectiveness for search models.

The problems of evaluation and of defining users are generalized by proposing that they are related to the need for an unified framework for defining arbitrary search concepts, search systems, user models, and evaluation strategies. It is argued that this framework depends on a re-interpretation of the concept of search accommodating the increasingly embedded and implicit nature of search on modern operating systems, internet and networks. The re-interpretation of the concept of search is approached by considering a generalization of the concept of ostensive retrieval producing definitions of search, information need, user and system that (formally) accommodates the perception of search as an abstract process that can be physical and/or computational.

The feasibility of both the mathematical formalism and physical conceptualizations of quantum theory (QT) are investigated for the purpose of modelling the this abstract search process as a physical process. Techniques for representing a search process by the Hilbert space formalism in QT are presented from which techniques are proposed for generating measures for effectiveness that combine static information such as term weights, and dynamically changing information such as probabilities of relevance. These techniques are used for deducing methods for modelling information need change. In mapping the 'macro level search' process to 'micro level physics' some generalizations were made to the use and interpretation of basic QT concepts such the wave function description of state and reversible evolution of states corresponding to the first and second postulates of quantum theory respectively. Several ways of expressing relevance (and other retrieval concepts) within the derived framework are proposed arguing that the increase in modelling power by use of QT provides effective ways to characterize this complex concept.

Mapping the mathematical formalism of search to that of quantum theory presented insightful perspectives about the nature of search. However, differences between the operational semantics of quantum theory and search restricted the usefulness of the mapping. In trying to resolve these semantic differences, a semi-formal framework was developed that is *mid-way* between a programmatic language, a state-based language resembling the way QT models states, and a process description language. By using this framework, this thesis attempts to intimately link the theory and practice of information retrieval and the evaluation of the retrieval process. The result is a novel, and useful way for formally discussing, modelling and evaluating search concepts, search systems and search processes.

Acknowledgements

The research for this thesis lead me on a significant trip in the land of knowledge among the landscape of subtleties of various fields, from the natural sciences and social sciences, to the computational and information sciences. It internally re-emphasized in the strongest possible way my interests, not in the form of any one discipline but as a formless curiosity that defies to be restricted by unnecessary boundaries. Due to the 'self-referential nature' of this topic as encapsulated by 'search is the perennial yearning to be found', the topic of research chosen led to interesting exploitations of the intellectual type that challenged how I approach learning and thinking in general.

I am deeply indebted to Professor Keith van Rijsbergen, describing whom merely as 'academic supervisor' would be grossly inaccurate, who speculated the intellectual trip I was to take when starting the research and facilitated it (often from behind the scenes) from the intellectual, social and administrative fronts. From our first discussion from which I came away understanding the concept of quantum theory as being 'between' computation (especially retrieval) and cognition I was hooked to this research and trying to understand this 'between-ness'. That first discussion re-sparked my interest in research into computing and introduced an interest in information retrieval and kept me determined to do formal research. I am also indebted to my other supervisor and academic mentor Dr. Joemon Jose who provided the needed 'view from the bridge' that kept me concerned about the practical side of the research, and consistently reminded me about the the views of the research community. How can one possibly pay the debts owed to teachers who facilitate the development of ones understandings about the world that deeply influence one for life?

I would especially like to acknowledge the IR group (past and present) at Glasgow especially Ryen, Sumitha, Leif, Christina, Vassilis, Azreen, Ali, and Alvaro, for entertaining yet linearly challenged discussions which helped me broaden my ideas on this subject. Further, I would additionally like to generously acknowledge Guido, Frank, and Yashar for critically reading parts of the thesis and/or informative talk. I also benefited much from technical discussions with Dawei Song, Mounia Lalmas, Peter Bruza, Matt Leifer and Alex Wilce.

Of tremendous benefit were the seminars on Information Science at the School of Informa-

tion at the University of California at Berkeley and inspiring discussions with Professors' William S. Cooper, M. E. Maron, Ray Larson, Michael Buckland, Lotfi Zadeh (at the Berkeley Institute for Soft Computing) at Berkeley and Stanley Peters at the University of Stanford (Center for the Study of Language and Information).

Of particular benefit were the extensive chats on wide ranging topics with Michael, Jahan Zeb Waheed, Amar, Hamza, and many other friends who provided necessary distractions from work. I would like to especially acknowledge Sohail for the concentrated study sessions in the write-up period.

I would like to thank the Engineering and Physical Sciences Research Council for partially funding my work and the Royal Society of Edinburgh for funding an inspirational research visit.

I would especially like to acknowledge my parents for persistent encouragement and support in various forms.

And what of the *validity* of gratitude if not directed at the One from whom originates all aid, to whom be All Praises.

Papilio tempestae (Quantum Weather Butterfly)

Contents

1 Introduction **1**

1.1 Introduction . . . 2

1.2 Motivation . . . 3

  1.2.1 Introduction . . . 3

  1.2.2 Sociological Motivations . . . 4

  1.2.3 Motivation from considering Search as Education . . . 7

  1.2.4 Cognition Modelling Motivations . . . 8

  1.2.5 Retrieval Modelling Motivations . . . 10

  1.2.6 Motivations due to the requirement of a *'Search Science'* . . . 11

  1.2.7 Motivations due to the requirement of a new definition of Search . 12

1.3 Thesis Statement . . . 13

1.4 Thesis Outline . . . 15

2 Background **18**

2.1 Introduction . . . 19

2.2 Traditional IR . . . 20

  2.2.1 Relevance and Relevance Feedback . . . 25

  2.2.2 Evaluation . . . 27

2.3 Ostensive Retrieval . . . 28

2.4 Cognition . . . 31

  2.4.1 Introduction . . . 31

  2.4.2 Structure of Cognition . . . 33

  2.4.3 States and State Change . . . 35

  2.4.4 Purpose of Cognitive Agents . . . 37

  2.4.5 Ostensive Model as Cognitive Agent . . . 38

  2.4.6 Quantum Theory for Cognition . . . 40

  2.4.7 Summary . . . 41

2.5 Defining the Search System . . . 42

2.6 Discussion . . . 44

  2.6.1 Traditional and Ostensive Retrieval . . . 44

2.6.2 The Evaluation Problem 46
2.6.3 Unified Retrieval Model 49
2.6.4 Method and Representation Spaces 51
2.6.5 Abstraction of Research Problems 54
2.7 Summary 56

3 On Reduction to the Higher Ostensive 57
3.1 Introduction 58
3.2 Notion of State 58
3.3 Agents, Knowables and Incompleteness 60
3.4 The Ostensive Philosophy and State change 64
3.4.1 The Higher Ostension 67
3.5 User's Self-Knowledge and Simulation 71
3.5.1 User Simulation 73
3.6 Non-Determinism and High-Level Search 77
3.7 What is Search? A Dialectic from First Principles 78
3.7.1 Scenario Approach 80
3.7.1.1 Ontological Approach 81
3.7.1.2 Dilemma due to Relevance Judgements 83
3.7.2 Approximate Users 84
3.7.3 Separation and Association of User and System 85
3.7.4 Resolving the Conceptual Problem 89
3.7.4.1 Quantum State Collapse and the Higher-Ostension 92
3.8 Summary 94

4 On Using Quantum Theory for Search 95
4.1 Introduction 96
4.2 Background 97
4.2.1 Introduction 97
4.2.2 The Hilbert space and Dirac Notation 98
4.2.2.1 Vectors and Operators 99
4.2.2.2 Inner Products, Eigenvectors and Eigenvalues 100

4.2.2.3 Hermitian Operators, Projectors . . . 101
4.2.2.4 Operator Functions . . . 102
4.2.3 Representing State and Evolution . . . 103
4.2.3.1 QT Postulates pertaining to State and State Evolution . 103
4.2.3.2 Superposition . . . 103
4.2.3.3 Convex Combinations and the Generation of Probabilities for Events . . . 104
4.3 Ways of Using QT . . . 106
4.3.1 Introduction . . . 106
4.3.2 Using QT vs. Using the Mathematics of QT . . . 107
4.3.2.1 Use of QT Axioms . . . 107
4.3.2.2 Using Mathematical formalism of QT in a Specific Way . 107
4.3.3 Discussion of Conceptual Approaches . . . 108
4.3.3.1 Introduction . . . 108
4.3.3.2 Discussion of Conceptual Approaches to modelling from QT . . . 108
4.3.4 A Technique for Modelling Simple Searches . . . 113
4.3.4.1 Introduction . . . 113
4.3.4.2 Relevance States . . . 114
4.3.4.3 Representation of Search paths . . . 114
4.3.4.4 Evolution of relevance states . . . 118
4.3.4.5 Visualising in Polar Co-ordinates . . . 123
4.3.4.6 Comparing search paths . . . 125
4.3.4.7 Interpretation of ν_0 . . . 126
4.3.4.8 Discussion of Utility Measures . . . 129
4.3.5 Measures over States . . . 131
4.3.5.1 Mixed Static/Dynamic Measures . . . 132
4.3.5.2 Measures as Functions of Other Measures . . . 132
4.3.5.3 Trace Measures . . . 133
4.3.5.4 Incorporating Context into Measures . . . 136
4.3.6 Change of State and Dynamics . . . 136
4.3.7 Simulation Design . . . 137

4.3.7.1 Exploiting Symmetry . . . 138
4.3.7.2 Morphisms . . . 140
4.4 Summary . . . 141
5 On Search as a Communication Process 143
5.1 Introduction . . . 144
5.2 Visualisation in terms of stacks . . . 145
5.2.1 Modelling of Diverse Information Seeking Scenarios . . . 148
5.2.2 Topology of visualisation . . . 148
5.3 Search Scenario Statics, States and Ontology . . . 151
5.3.1 Researcher Types . . . 151
5.3.2 States, substates and statements . . . 154
5.3.2.1 The Structure of the General Search State S_{Π} . . . 154
5.3.2.2 State structure . . . 156
5.4 Dynamics, Flows and Epistemology . . . 159
5.4.1 Introduction . . . 159
5.4.2 Semantics . . . 161
5.4.2.1 Interaction . . . 163
5.4.2.2 The Non-Realism of the AU's . . . 164
5.4.3 Shared Knowledge . . . 164
5.4.4 Measures on Sets . . . 166
5.4.5 Flow Types . . . 168
5.4.6 Structure of Semantics . . . 170
5.4.6.1 Constant properties for semantics . . . 171
5.4.6.2 State and Energy . . . 172
5.4.7 Using Flows in a Scenario . . . 173
5.4.7.1 Picking States . . . 173
5.4.7.2 Relation to Higher Ostension . . . 177
5.4.8 Convergence . . . 178
5.4.8.1 Simulation and Convergence . . . 183
5.4.9 Higher-level Semantics . . . 184
5.5 Languages and Translation . . . 187

    5.5.1 Introduction 187
    5.5.2 Language and Cognition 188
        5.5.2.1 Structure of $S_{cog} \subset S_{user} \subset S_{\Pi}$ 189
        5.5.2.2 Limit of Languages 191
        5.5.2.3 Derivation of Languages 194
        5.5.2.4 Cognition Modelling issues 194
    5.5.3 Translating between Frameworks 196
        5.5.3.1 Introduction 196
        5.5.3.2 Observation ($\lfloor$) and translation 196
        5.5.3.3 Difficulties with R_{KVR-QT} 199
        5.5.3.4 Simulation Techniques 200
5.6 Summary 202

6 Applications and Discussion 204

6.1 Introduction 205
6.2 Structured Document modelling and retrieval 205
6.3 Translating to QT and Conforming to R_{QT} 208
        6.3.0.5 Strategy for using R_{thesis} 211
6.4 Use of R_{QT} in light of R_{thesis} 212
    6.4.1 Utility Measures, Convergence and Exhaustive Analysis 212
    6.4.2 Developing Quantitative Models in R_{QT} 213
    6.4.3 Deducing Probabilities of Relevance 214
    6.4.4 Possible uses of the Measurement Method 215
    6.4.5 Formulating new search models 215
6.5 User Modelling 217
6.6 Evaluation Methods 221
6.7 Structural Comparison of Models 226
6.8 Discussion 227
6.9 Summary 233

7 Conclusions 234

7.1 Conclusions 235

7.1.1 Ostensive search represents any search . . . 235
7.1.2 Techniques of Analysis from Quantum Theory . . . 236
7.1.3 General Representation Techniques . . . 236
7.2 Future Work . . . 238

A Encoding **241**
A.1 Probabilities into Angles . . . 241
A.2 Document-Level Utility . . . 241

B Representing terms as Signals **245**
B.1 Introduction . . . 245
B.2 Encoding Terms . . . 246

C Notes on Evolution **249**
C.1 Hamiltonians, Super-selection and Ostension . . . 249
C.2 Using Histories . . . 252

List of Tables

6.1 Summary of similarities between concepts in R_{QT}, R_{CS}, R_{trad} and R_{thesis} . 232

List of Figures

1.1 The social gap in automation 4
1.2 The pre-query stages of search 11

2.1 Data flow in the Laboratory model 21
2.2 The pre-query stages of search 22
2.3 Simple physical model with underlying theory 24
2.4 The ostensive path 30
2.5 System imitates human 32
2.6 Reasoning Process 33
2.7 Association of Concepts 34
2.8 Uncertain logic from associations 34
2.9 The OM as a cognitive computation device 38
2.10 Types of State Change in OM 39
2.11 Knowledge State Flux in the Ostensive Model 45
2.12 Method and Representation spaces of search components 52
2.13 Mapping of Method and Representation Spaces 53
2.14 Dependence of Research Problems 55

3.1 Theory as an Abstract Device 59
3.2 Knowledge of Agents 61
3.3 Re*search 63
3.4 Researcher Perspective 65
3.5 System Perspective 66
3.6 User Perspective 66
3.7 Research Oracle's Perspective 69
3.8 External Oracle's interaction with Search 69
3.9 Researcher perspective in Search Simulation 70
3.10 Researcher Perspective in User Modelling 70
3.11 User Modelling 76
3.12 Information Need Path 77

3.13 Summary of Relationships between Search and High-Level Search 78

3.14 Model for an Approximate User 84

3.15 QT Researcher's Perspective of Reality 93

3.16 An object-oriented perspective of the relationship between search and high-level search where A :: B indicates that concept A is an 'instance' of concept B 93

4.1 The Wave method R_w and Observation/Measurement method R_m for modelling phenomena 110

4.2 Geometric Metaphor for relationship between phenomena, measurement and observation approaches 112

4.3 The ostensive browsing process represented by a graph 115

4.4 A graphical representation of the result of adding a set of complex numbers in polar-co-ordinate form that represents the result of operations of the relevance state 124

4.5 Inner product as Potential Difference: An analogy to aid interpretation of inner product between states 128

5.1 A Stack Model 147

5.2 Model of an Evaluation 148

5.3 Two different visualisations of the same agent, showing components and two types of internal flows. The general representation on the left and stack on the right 150

5.4 A non-linear stack of six components each with three or five adjacent components 151

5.5 Substates corresponding to components of a system displayed as a stack . 161

5.6 A diagram of the research & refinement process showing the initial concepts and the goal concepts, which are yet to be tackled 171

5.7 Visual similarities between ostensive retrieval interface and general state change 174

5.8 Flow with explicit parameters 177

5.9 A simple search stack 183

5.10 Association of Concepts 191

5.11 Observing traditional information need transformation in terms of languages 192

5.12 Uncertain logic from associations 196

6.1 A document structured in three ways, by sentences, sections and topics, with flows between items . . . 206
6.2 A flow between elements at different semantic levels, a word and a section in a document in this case . . . 207
6.3 A simple stack design . . . 210
6.4 A stack design that considers high-level semantics for user modelling . . . 219
6.5 User's Self-knowledge denoted by a state observing itself through an evaluatory second role . . . 221
6.6 Evaluator as a separate agent that is an oracle relative to both simulated user and system . . . 221
6.7 A possible substate structure in simple descriptive representation for the information experience that is sought S_{IE} . . . 222
6.8 The perceptions explored in this chapter related with respect to observation operators $\lfloor$ or $\widetilde{\lfloor}$ . . . 228

List of Symbols

$\mathtt{RA}$	Abstract system or abstract retrieval agent	79
$\mathtt{AU}$	Abstract User (a special case of which is the approximate user)	84
$\mathcal{H}$	A Hilbert Space	97
$\lvert x\rangle$	A unit column vector or ket (in Dirac notation)	98
$\langle x\rvert$	A unit row vector or bra (in Dirac notation)	98
$\langle y\lvert x\rangle$	An inner product represented (in Dirac notation) as a multiplication between a bra and ket	98
$\overline{\langle y\lvert x\rangle}$	The complex conjugate of $\langle x\lvert y\rangle$	98
$A^\dagger$	Conjugate transpose of matrix A	98
$\mathcal{H}^\dagger$	The dual space of $\mathcal{H}$	99
$\mathcal{I}$	The set of linear functionals	99
L_v	Space of Linear Operators on $\mathcal{H}$	99
$\langle y\lvert A\rvert x\rangle$	Transformed operator application, can be thought of as the functional $\langle y\rvert A$ applied to $\lvert x\rangle$	100
$\lvert x\rangle\langle y\rvert$	The outer product between two vectors producing a matrix	100
$\mathrm{tr}(A)$	The trace function of A, its value is the sum of diagonal entries of A	102
$\lvert i\rangle$	Denotes the basis for a Hilbert space	102
$\lvert\psi\rangle$	A unit vector representing the state of a system	103
R_w	The researcher (or researcher agent, see Chapter 3) using the wave or phenomena perspective	109
R_m	The researcher or researcher agent using the measurement or observational perspective	109
$R_{w-thesis}$	The researcher or researcher agent (see Chapter 3) interpreting the concepts of this thesis according to the wave or phenomena perspective	111
$R_{m-system}$	The researcher or researcher agent using the measurement or observational perspective to model a (retrieval or general) system	111
$R_{w-system}$	The researcher or researcher agent interpreting a (retrieval or general) system according to the wave or phenomena perspective	111

$\mathbf{R_{w-user}}$	The researcher (typically representing a system in this case) interpreting the user according to the wave or phenomena perspective	111
$R_{observer}$	The researcher or research agent at a higher-level than R_m and R_w in that it is able to observe them and compare between them. It represents an abstract device or theory such as Quantum Theory through which one observes phenomena.	111
$R_{external}$	The researcher or research agent, of the type of the external research oracle in Chapter 3 that is at a higher level that $R_{observer}$ which it is able to observe. It is therefore able to compare between theories (abstract devices) which an $R_{observer}$ such at Quantum Theory is unable to do in an automatic way. An example of $R_{external}$ is this thesis.	111
$\|\Omega\rangle$	Information Need State, representing user cognition	114
$\|\psi\rangle$	Relevance State, representing system "cognition"	114
$\overrightarrow{OG}$	Ostensive graph, represents user's interaction paths according to the ostensive model	114
δ	Degree of the ostensive graph, denoting the number of documents suggested to the user following a document selection	115
T	The set of terms or an operator on $\mathcal{H}$ depending on context of use	116
${}_{t}p_i$	Complex number representation of the probability of relevance of item i	116
${}_{t}\chi$	The set of all items with which the user can interact with at time t in the ostensive interface	117
U	The unitary operator on the Hilbert space, used to represent the evolution of states	118
ω_j	The j^{th} eigenvalue of a unitary matrix, $\omega_j \in \mathbb{C}$	119
${}^{d}_{t+1}U$	The evolution operator at time $t+1$ represented in word/term basis	119
${}^{d\rightarrow w}_{t+1}U$	The evolution operator at time $t+1$ represented first in document basis then re-represented (basis changed) in word/term basis	120
${}^{w\rightarrow d}_{t+1}U$	The evolution operator at time $t+1$ represented first in term basis then re-represented (basis changed) in a document basis	120
ω_i	Complex number entries in the matrix representation of the Unitary operator U	120
${}_{t}\Theta_i$	The resultant dynamic weight at time t for term or document i	121

$\nu(\psi_a, \psi_b)$	The Utility function that calculates the inner product between states or documents	123
ν_0	General form of the Utility function	125
$\nu_{i,j}$	A summand of the utility function between documents i and j representing the inner product multiplied by the difference in probabilities of relevance of the respective documents	126
$Re(X), Im(X)$	Functions that return the real and imaginary parts of complex number $X \in \mathbb{C}]$	129
ν_1, ν_2	Forms of Utility function for special case relationships between states and/or documents	129
${}^d\nu(\psi_1, \psi_2)$	Utility function between states which are represented in a document basis	130
D, D_i	Alternative symbol for $\|d_i\rangle\langle d_i\|$ that denotes a matrix (and symbolic representation) of a document	131
$f_{\textsf{stat}}, {}^{\textsf{stat}}f$	A measure on the set of documents, states or terms employing static information	131
$f_{\textsf{dyn}}, {}^{\textsf{dyn}}f$	A measure over the set of documents, states or terms employing dynamic information	132
$f_{\textsf{mix}}, {}^{\textsf{mix}}f$	A measure over states, documents or terms employing static and dynamic information	132
$\|_0d\rangle\langle_0d\|$	The initial diagonal matrix in term basis holding static weights of terms in document d	133
$dyn(D)$	An operation on document matrix D which results in the removal of its static weights such that the result is a diagonal matrix with only dynamic weights (typically probabilities) remaining	133
${}^w\nu\,(D)$	The Utility of a single document expressed in term basis	134
${}^w\nu(D)$	Term-level utility measure over a single document	134
${}^w\nu(\psi)$	Term-level utility measure over a single state	134
${}^m\nu({}^w\psi_a, {}^d\psi_b)$	Mixed-Level Utility function between two states	135
$T_{d\rightarrow w}$	The projection (re-representation) that denotes the change of basis from the document space (in document basis) to term space where term vectors form an orthonormal basis	135
$\nu(D), \nu(\psi)$	Single element utility function for a document/state	135
ψ_o	The state with the most utility among a set of states; Optimal state	135

D_o	The document with the most utility among a set of document; Optimal document	135
$G(U, \times_{nxn})$	A group on the set of unitary operators	138
${}_{t_1,t_2}X \subset U$	A subgroup representing the set of changes to a (search) state between time periods t_1 and t_2	139
${}_tU_{a,b,c}$	A set of sets, containing the the sets of changes in terms indexed a, b and c	139
${}_t\dot{U}$	A group of changes all of which are simply derived from the change U; a cyclic group of changes	139
ϕ	A map between two groups (e.g. of user behaviours) that preserves the groups properties; can be used to indicate the change in the behaviour of a process	140
R	A researcher, abstract device (see Section 3.1), perspective and/or framework by which a process can be observed	151
R^*	The optimal researcher, a hypothetical entity that knows all statements and their truth values for any search	151
$\mathcal{R}$	The set of researchers, perspectives, frameworks and/or abstract devices in which R^* is the (hypothetical) largest element	151
$R_1 \preceq R_2$	A binary order on the set of researchers indicating that R_2 knows at least as much as R_1 for all processes that can be observed (since no process has been specified in the relation)	151
$R_1(\Pi) \preceq R_2(\Pi)$	Denotes that R_2 knows at least as much as R_1 throughout the entire (search) process Π	151
$K(R), K(R(\Pi))$	The knowledge potential function. It outputs from the (hypothetical) set of all possible statements, a set of statements and corresponding truth values pertaining to R.	152
Π^*	The set of all processes, all of which are *known* to R^*	152
R^*_{sim}	The optimal researcher for simulated search scenarios	152
$R_{res.orac} \prec R^*_{sim}$	The research oracle perspective, which knows all possible states a search can go to at some point in time without knowing which actual state it will go to; thus note that $R_{res.orac} \prec R^*_{sim}$	152
R_{vect}	The traditional vector space framework	153
R_{log}	The traditional logical framework	153
R_{prob}	The traditional probabilistic framework	153
R_{KVR-QT}	The QT inspired framework from [vR00]	153

$R_a \lfloor_{\Pi} R_b$	Denotes the process of observation/research/analysis and therefore (high-level) search, in which R_a is the observer and R_b the observed (see Chapter 3)	153
R_{trad}	The research approach that is the highest in the partially ordered set of all traditional approaches to IR (thus at the least $[R_{vect}, R_{log}, R_{prob}] \preceq R_{trad}$)	153
R_{thesis}	A composite researcher encompassing all the perceptions, models, understanding and theoretical devices presented by this thesis	154
${}_tS_{\Pi}$	The superstate (or most general state) of the process Π at time t, it refers to all knowledge a researcher has of the process at that time	154
$\mathcal{L}(S)$	The set of all statements pertaining to the representation (and general characteristics about expression) of S	155
${}_tS_{\Pi} = \bigcup_i {}_tS_i$	The system state as a composition of substates	156
$R_{QT}, R_{QT}(\Pi)$	A QT perspective (in general, of any process) and the QT perspective of Π	157
s_a, Σ_{Π}	A statement and the set of all statements of a process Π which is a partial output of $K(\Pi)$	158
$T(X)$	A translation of a statement into another statement (usually in different language or representation, or equivalently a map of one framework (or framework component) into another framework (or framework component))	158
R_{IR}	The IR perspective of a process and/or IR framework	158
$A <: B$	A is a subtype of the type (associated with) B	159
γ_{init}	The initial external flow (influence) that initiates changes in a process	160
$S_{L_1} = \bigcup^{L_1} S_i$	The substates of layer L_1 as represented by a stack in the stack diagram	160
${}^{A}_{t}S_i$	State for sub-component i of component (or 'layer') A at time t	160
Γ_E	Set of external flows	161
Γ_I	Set of internal flows	161
${}^{\Pi}\Gamma$	Set of all flows associated with process Π	161
Γ_{EXP}	Set of flows that pertain to the expression subprocess of an agent	161
${}^{\Pi}\mathbf{P}$	The set of all paths in Π	162
$\wp\left({}^{\Pi}\Gamma\right)$	The power set, or set of all sets, of flows for Π	162

${}^{\Pi}\widetilde{\mathbf{P}}$	Set of all possible ordered paths in a process Π	162
$\widetilde{p}$	An ordered set of flows, or ordered path	162
${}^{\Pi}\widetilde{\mathbf{PS}}$	An ordered set of ordered set of flows or an order set of ordered paths	163
$\mathcal{I}\left({}^{\Pi}P_1\right)$	The (traditional) information content of the path P_1 in Π	166
$\mathcal{FT}$	The set of all flow types within all agents for Π, and the flow supertype	169
$\mathcal{FT}_I$	Set of all internal flow types, and internal flow supertype	169
$\mathcal{FT}_{INT}$	Set of all interpretation flows, and interpretation flow supertype	169
$\mathcal{FT}_{EXP}$	Set of all expression flows, and expression flow supertype	169
$\mathcal{FT}_{\mathtt{AU},(i,i)}$	The flow type pertaining to the internal flows of component i of an abstract user	169
$\mathcal{ET}$	The energy supertype of a process	170
$a \subset^* b$	A relation indicating that a is either a subset of b or in a subset whose *eventual superset* is b	172
$\mathcal{ST}$	The State supertype	173
$K(S_\gamma)$	The knowledge pertaining to the flow γ	175
γ_ω	Flow with explicit parameter ω where ω <:S	176
$C : (S, S) \mapsto S$	Convergence function, a general schema for several types of functions	179
$T_{English}(X)$	Translation of a state referred to by X into natural language (English) form	182
Pick	A function (and subprocess of Π) for selecting the next state in a set of possible future states corresponding to the concept of higher ostension in Chapter 3	185
$\overline{\gamma_x}$	Flow going in the opposite direction to γ_x, usually used to indicate the inverse (undoing) of the change represented by γ_x	187
$\mathcal{L}_\infty$	The richest (hypothetical) language able to fully express human cognition	190
$\mathcal{L}_N$	The natural language	190
$\mathcal{L}_P$	Hypothetical description language for perceptions, perceptual language	190
$\mathcal{L}_G$	The language of R_{thesis}, specifically of Chapter 5	195
$R_{English}$	A descriptive research perspective expressed in natural language, specifically in English	210

$E_2 \perp E_1$	Operator or space E_2 is perpendicular to E_2, used to denote the distinctness in the types of two properties	215
$S_{IE} \subset S_{cog}$	The state representing the information experience of an agent; it is included in the 'cognition' of the agent	219
RU	A real user	225
UM	A user model	225
R_{HHIR}	The HHIR research perspective	227
$R_{soc.sci}$	The sociological and social scientific research perspective	227
λ	A parameter for scaling probabilities of relevance in a complex number encoding	241
ϵ	Approximation quantity for calculation of utility	242
$\mathbf{H}, \mathbf{H}(t)$	A Hermitian operator known as the Hamiltonian in time-independent and time-dependent form respectively	250
$dim(\mathcal{H})$	A function that returns a positive integer denoting the dimension of space $\mathcal{H}$	250
$Y_1 \odot Y_2$	The space composed by the tensor product of Y_1 with Y_2	252

—Let the beauty of what you love be what you do.

Jalaluddin Rumi (1207-1273)

1

Introduction

1.1 Introduction

Information retrieval (IR) is the the field of research investigating the searching of information in documents, searching for documents, searching for *information about* documents, or searching within databases, whether stand-alone or networked by hyper-links such as the internet. Types of data searched include text, audio, video, images or other complex data types such as programs. An IR system is commonly understood as that which deals with the relationship between objects and queries. Queries are formal statements of *information needs* addressed to an IR system by the user. The object is an entity which stores information in a data set, known as a document. User queries are matched to documents stored in document collection. Often the documents themselves are not kept or stored directly in the IR system, but are instead represented in the system by their pointers.

Automated information retrieval systems were initially used to manage information explosion in scientific literature in the last few decades. The value of information is directly related to its ability to be located and used effectively, search engines thereby form a crucial component in the research and understanding of modern times. For something so crucial, IR can be a confusing area of study. Firstly, there is an underlying difficulty with the very definition of IR as there exist the adjacent fields of data, document and text retrieval [Bla84]; knowledge, information and data management [Bla02]; information seeking[IJ05, Cas06, Mar95], information science [Wil97, VV87] and others with their own bodies of literature, theory and technologies which are deeply related to IR and each other to the point where the boundaries are unclear. Secondly, IR is a broad interdisciplinary field, that draws upon secondary fields such as cognitive science, linguistics, computer science, library science and it does so in a loosely organized fashion. It is tempting to refer to this conjunction of diverse areas as *"search science"*. However, due to the presence of ad-hoc techniques used to perform experimentation in IR and the absence of a (general) formal language for definition of IR concepts, components and results, it cannot be called a science[1]. Furthermore, there are no specific definitions of search. With the abundance of methods available for finding information, whether through computer applications, libraries and librarians, a combination thereof, or otherwise, a formal definition would need

---

[1] This is further elaborated in the next chapter.

to accommodate a process far more complex than that of traditional web-based querying through systems like Google. The lack of a general formal specification method for search processes, IR research, and the absence of a strict scientific method underpinning it has posed major barriers to future development and usefulness of research in the field [AvRJ05].

This thesis addresses the conceptual, methodological and theoretical foundations of information retrieval with the intention of creating standard principles for understanding search in its various forms and hoping to thereby enhance research methodology. Borrowing notions from the mathematical formalisms, operational methods and interpretational mechanisms of Quantum Theory (QT), this work aims to show that the conceptual ambiguities underlying current research methods are responsible for many of the research problems. Alternative ways of understanding search are proposed with corresponding methods of conceptualization, some allowing mathematical formalization. The next section discusses the broad motivations for addressing the foundations of search.

1.2 Motivation

1.2.1 Introduction

In the last decade and half the process of searching has been characterized by the interaction between a user and the textbox and search button pair followed by a browsing process. Compared to human-human interaction where the participants use multiple senses, a complex language and possess advanced reasoning systems, *human-computer information retrieval* (HCIR) is not just primitive but quite unnatural for the human and also unnatural for the computer unless it can accurately model the user or more specifically the users information need (IN). The next few sections consider that human-human interaction, specifically *human-human information retrieval* (HHIR), requires to be understood from a sociological perspective and that human-computer interaction, in particular HCIR (searching) takes on a special sociological role in current times and encourages a particular social perspective in a searcher searching over long periods of time. This sociological perspective is known as the *sociological imagination* [Mil59].

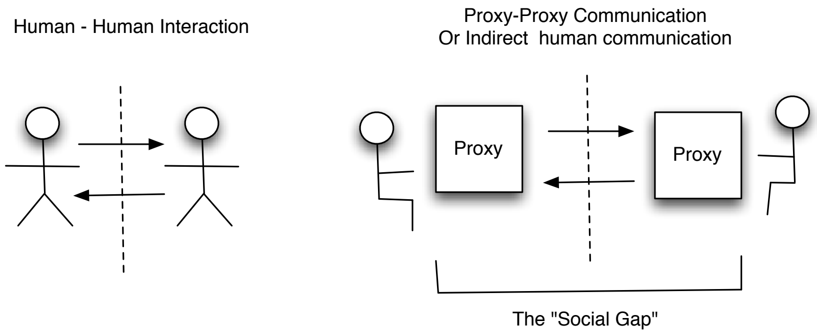


Figure 1.1. The social gap in automation

The the next few sections discuss the motivation from a sociological, educational and cognition modelling point of view. The motivations from the retrieval research point of view are then addressed in Section 1.2.5 and Section 1.2.6. The aim is to show that at the root of the motivations from these different perspectives is a motivation pertaining to the re-definition of search as a concept that better accommodates the related concepts from the areas.

1.2.2 Sociological Motivations

One of the results of automation of services through use of computation is the decrease of human-human interaction in the areas that would otherwise be required to provide and consume these services. Instead, computation has replaced several avenues of interaction between people and as a result human-human interaction has been substituted by the interaction of proxy devices that represent people. The proxy creates a *social gap* between people by replacing the function of interaction with abstract representations by information about that which leads to the gap[2], as depicted in Fig. 1.1.

Some social gap is inevitable for automation, yet automation could increase human interaction in other areas; for example on the internet, the indirect interaction is increased through proxies due to e-mail, social networking, and other general online communication. Interaction between people and entities that are already proxies, such as books and written

---

[2] The social gap is a generalization of the concept of the semantic gap in IR which is the difference between the form of a retrievable object and its meaning. As here social gap refers not to difference in form and meaning of a retrieval object but a general proxy object that, prior to proxies, would be a human or human interactions.

knowledge increase through automation as made apparent by the increasing use of the internet for academic purposes. However, social interaction teaches more than just factual information, it facilitates the development of identity through social analysis of other people. How then is a person's identity defined by their daily *information experiences*[3] in place of social interactions and caused by the social gap?

The *sociological imagination* [Mil59], meaning the lucid awareness of the relationship between personal experience and the wider society, addresses issues located on the crossroads between one's private identity and its relation to the public sphere. It is proposed that the web, social networking applications and other networked applications form the 'public space'[4] for computer users, but in this case the concepts of this public space are developed in one's private space. In this sense a computer user is still concerned with private issues but the public issues become perceived in a private fashion due to the way this public information is accessed (in the privacy of a home environment) and conceptualized, yet most people may remain focused on private issues without realizing the social reality in which the issues are embedded. The sociological perspective helps to see general social patterns in the behaviour of individuals and offers insights about the social world that extend beyond explanations relying on individual personalities. Essential to this sociological perspective is the sociological imagination, which means going beyond the individual and understanding how social structures shape individuals and their action. In the web which consists of proxies of multiple people, the sociological imagination is encouraged by nature of the internet with respect to the ideas that one can form about people from their proxies, thus the sociological imagination applies at the level of proxies[5].

As more people are represented by their proxies on the web, and with increasingly sophisticated ways to represent aspects of their personality, the individuals 'web life' becomes more structured, to what extent can one take a users proxy representation as a reflection of their real selves, and to what extent can sociological investigations be conducted with proxy information? The extent depends on the relationship between the proxy and the person represented, and also in general between man and machine[6]. In order to under-

---

[3] The concept of an information experience is further addressed in Chapter 3.

[4] The public space forms part of one's proxy, it contains the information about one that is exposed to others.

[5] The social gap otherwise restricts the physically motivated sociological imagination.

[6] The general case would address theoretical limits whereas the a measure of similarity between a proxy

stand the extent to which one's proxy information represents themselves from a sociological perspective one needs to decide what it is about one that is embedded in proxy form, and more importantly, what the overall purpose of automation is or should be. In the most basic level a human engages in the framework of automation that requires them to give proxy information in order to find out information[7]. Thus it is said that the impetus of the relationship between man and machine, or the underlying cause, is search. The interaction between man and machine, where man is directly requesting information from a system, as opposed to using it as a proxy to ask another human, is a form of socializing from which one can understand things about the user, hence as search behaviour is sociological activity, it is proposed that it *induces a sociological imagination*.

Moreover this thesis proposes that search is a *primitive*[8] with respect to it being a function, tool, action and activity, that underlies the formation of identity through the use of a computer. Search takes on the role of a special member in the 'society'[9] by which one forms identity, its use reflects the properties of the user, specifically it is used to express the sociological perspective of its users. This is easier to see if one does not perceive search as an isolated activity that is a minor part of computer usage. As search is not simply the use of simple web-based tools, it is also the user interacting with documents, selecting programs, saving documents or any activity as all activities in a way, however minor, affect one's idea of themselves and the world around. With respect to developing identity through sociological imagination it can be said that the user is searching for what in hindsight would be their future identities or future selves described in terms of their 'interaction proxies' that contain instances of general computer usage, interactions with social applications on the web, computer-usage influenced psychology, and other effects that occur in the human social realm through interaction with the artificial world. These are issues that are both associated to the entity 'in front of screen' as well as being on a theoretical level 'on top' of the technical components 'behind the screen' in the same way

---

and the person represented would be a measure of *practical effectiveness*.

[7] The giving of proxy information i.e. making a webpage, is understood in this thesis as an *information experience* in the sense that one is '*finding out about* the experience of making the website', this is a generalization of what it means to 'find'/'search' and is addressed in Chapter 3.

[8] In that it is not derived or developed from any other concept or process.

[9] Society refers to combination of people and their proxies. Search is a special proxy in this society that brokers interactions between a human and other proxies, in that it is the cause of interactions and also the nature of the interaction itself, see the concept of *higher ostension* in Chapter 3 for a generalization of this nature of search.

as human-computer interaction studies is on top of assembly code, classical physics is 'on top' of Quantum Physics[10] or chemistry is on top of physics even though not every item on top can be fully understood in terms of that which constitutes the lower layers.

As search is the impetus for interaction, the primitive sociological activity in HCIR/HHIR, and as proposed in Chapter 3 search *is* interaction and any interaction is search, thus the study of search is not only broad as is already known but could offer insights to other fields like sociology, thus in this way, search is important at the philosophical level to other disciplines[11]. It is hoped that this thesis will suggest ways to re-investigate the traditional concepts in IR and that of defining search in order to extend it so it can be used as a theoretical tool to address computer use on a social, technical and philosophical level.

1.2.3 Motivation from considering Search as Education

In terms of human-human interaction, education in the modern sense can be understood as a set of processes involving institutions (a set of proxies/people) and individual (proxies/people). The institution of education started locally in individual homes, then moved to the local community in the form of a school, later still an interesting phenomena became visible among many institutions as characterized by assembly line production strategy: the use of humans in an industrial process in the interest of 'automation' began to resemble the organizational structure of a machine-only process. The mass-production strategy influenced the education establishment; yet, although there is a reduced connection between individuals (as emphasized by the social gap) it certainly exists at the least within the bounds of social protocol within an institution[12]. This connection is most apparent between students and their educators as assumed in Chapter 2, where it is assumed that the HHIR in the educational context between academics represents the most effective or optimal search. In particular the educational context that is to be modelled for investigation of HHIR has changed over the years by integrating automated components. The

---

[10] This is since classical physics is at the macro level and quantum physics pertains to that which constitutes the macro level.

[11] Search as a practical tool is of general use, but it is proposed that search can be useful as a speculative philosophy for understanding among other things the mind-machine and mind-matter relationship, where speculative philosophy is understood as *"the endeavour to frame a coherent, logical, necessary system of general ideas in terms of which every element of our experience can be interpreted"* from p.4 of [Whi29].

[12] An example of social protocol is professionalism at work.

idea of proxies for education is not new, an educational culture based on automated, and 'ad-hoc' learning that is a reality in our time, was conceptualized as an 'educational web' in [Ill71]:

> Universal education through schooling is not feasible. It would be no more feasible if it were attempted by means of alternative institutions built on the style of present schools. Neither new attitudes of teachers toward their pupils nor the proliferation of educational hardware or software (in classroom or bedroom), nor finally the attempt to expand the pedagogue's responsibility until it engulfs his pupils' lifetimes will deliver universal education. The current search for new educational funnels must be reversed into the search for their institutional inverse: *educational webs which heighten the opportunity for each one to transform each moment of his living* into one of learning, sharing, and caring. *Deschooling Society, Ivan Illich* [Ill71]

In Illich's view, education in the form of educational webs[13] better accommodates the way people learn. The use of a structured information space such as the web with sub-webs corresponding to forums, blogs, and learning communities correspond to educational webs. Inductively one can say that people have formed on the web what is, in a restricted way, the effective learning environment that had been indicated by Illich. Further, this learning environment is one in which there is an *enforced social gap* that is expected to have a positive overall effect.

1.2.4 Cognition Modelling Motivations

Computational search can be characterized as the problem of forming a query that describes the user's information need (IN) and subsequently associating data entities to this need. One of the problems intrinsic to traditional IR is the query formulation problem (QFP) [tHPvdW96, URJ03] which refers to the difficulty on the users part in deciding what they are actually searching for. Query formulation is a difficult task for users as not only is their

---

[13] Which applies specifically to sub-internets such as wikipedia but in general the internet is clearly an ad-hoc educational web.

IN often vague but it is changing as they interact with the retrieval system. Why is writing a query unnatural or difficult, and how should dynamically changing needs be expressed and accommodated by a search engine?

Even with a well defined IN most users struggle to comprehend the relationship between their query and the system's response. This lack of understanding of how the retrieval system works limits the efficiency and effectiveness of search as the user fails to understand the system. What should then be the way in which a system is used, ought it to be as one with simple functionality[14] or with a set of particular types of search and interactions, what is the optimally useful tool in this regard? The problem characterized by these questions is described as the system understanding problem (SUP) which refers to the users lack of understanding of a search engine that hinders effective searching. The SUP and QFP are further elaborated in the next chapter.

It is not suggested that users should be expert searchers, or researchers of IR that know the technicalities of the query to document matching process but that the dialogue between user and system be more comparable to that between two users as in human-human information retrieval. In the HHIR context, it is the properties such as *behaviour* and *personality*, and the level of consistency of such properties that characterize the framework for communication and directly influence how agents in HHIR answer questions. Traditional IR systems are very artificial in comparison as they do not attempt to address such concepts directly. Should HCIR be more like HHIR? In HHIR, personality, behaviour and other factors, contribute to an overall framework for question answering, in HCIR the setup in a computer is artificial. What are then the necessary steps in the context of modelling that require to be addressed to make HCIR closer to HHIR. In particular, cognition modelling (as in [GÖ0] for example) presents a model for HHIR that is more realistic than traditional ways to model users in HCIR[15].

---

[14] Search and click interface of the original Google system.

[15] The idea of using QT as a language for cognitive modelling is discussed in Section 2.4.

1.2.5 Retrieval Modelling Motivations

One of the key reasons for delving into formalizations for retrieval is for addressing the problem of modelling changing information need, in particular, for modelling ostensive retrieval. Ostensive Information Retrieval (OIR) [Cam00] mainly aims to address vague and changing IN. It approaches the problem by defining user-system interaction so as to not require explicit query formulation. Change is accounted for treating certain interactions as more important than others. OIR is relatively new and has been shown to be effective for search, especially for image search. However, it is currently ad-hoc by nature as there exists no theoretically sound justifications for the specifics of its query-less interface or for the way it interprets IN change. Instead of requiring 'artificial' communication of the user by reducing the interpretation of their IN to a short phrase, OIR is an attempt to recover details of IN inevitably lost in the reduction by introducing an interaction method assumed to be more natural with respect to cognitive processes.

The idea of querying/feedback by *ostension* developed in OIR can be seen as a 'median' between explicit feedback/query formulation and implicit feedback as they are traditionally understood. OIR suggests using an *ostensive language* as (1) it improves the users understanding of how the system works, and (2) IN defined in such a language simplifies its interpretation for the system, meaning that changes in need become more transparent to the system. Ostensive IR by definition addresses the query formulation problem by removing the need to express a word based query. The Ostensive Model (OM) of developing information needs also recognizes the dynamic nature of information needs during a search process providing a simplistic model for interpreting change. It is limited in many ways relative to HHIR, one being its dependence on the binary probabilistic model [Fuh92], which means only simple 'relevant/not-relevant' type interactions by the user are captured, thus limiting user expression. The interactions are interpreted based on assumptions about information need change (a concept pertaining to cognition phenomena) by means of a simplistic model for cognitive state change.

Ostensive IR is a user centric search. A comprehensive investigation into OIR inevitably means addressing psychological and cognitive concepts or that some assumptions have to be made about these before user interactions and corresponding concepts of *expressing*

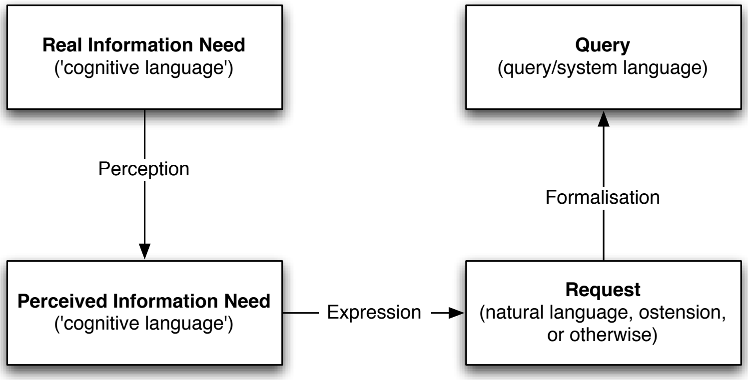


Figure 1.2. The pre-query stages of search

behaviour/personality can be adequately understood. In the past there have not been thorough, formal research presenting a cognitive user model for ostensive/interactive search. This thesis challenges the ad-hoc approach of the OM which limits the aims and potential achievable by OIR arguing that addressing the above issues, especially that of formal cognition modelling effectively, requires a broader theory for ostensive retrieval. The main motivations have been to investigate expansion and formalization of the OM: (1) the cognitive model of IN change, (2) the interaction model, (3) the uncertain inferences about IN made from interactions, (4) and how these affect retrieval. This thesis presents a framework in the language of quantum theory that addresses these issues, and the modelling of cognitive phenomena that pertain to information need change. In general, it was the advantage of OIR over traditional IR that motivated the investigation into the foundations of IR through a formal approach to OIR.

1.2.6 Motivations due to the requirement of a 'Search Science'

Recent work in [vR04] based on ideas borrowed from Quantum Theory (QT) have suggested methods of formalizing aspects of IR aiming toward a comprehensive theoretical basis in which a search process can be completely defined and reasoned about, and a scientific basis inspired by operational methods in QT. It was subsequently found that there is a potential for QT methods to play a wider role in resolving the above IR issues (of definition and lack of scientific method) than suggested. In addition, it was found that apart from the mathematical formalism of QT which offers methods for representation and analysis of IR concepts, the scientific method[16] and operational structure[17] are also very useful. Inspired by these peculiar connections and on attempting to apply these methods and *map* search to QT, it was found that search requires to be re-examined from a perspective quite different from how it is traditionally perceived (see [AvRJ05]) in order to deduce the feasibility, utility and method of the mapping. Thinking about search in this new way also suggests approaches for re-defining the concept of search [AvR07]. The overall goal for our research can be equated to being able to formally refer to IR as *"search science"* by establishing a specific definition of search and deducing scientific methods for the investigation of search, so it can be in all respects, a science.

This thesis addresses the conceptual foundations of information retrieval with the intention of creating standard principles and hoping to thereby enhance research methodology. Borrowing notions from the mathematical formalisms, operational methods and interpretational mechanisms of Quantum Theory (QT) this work aims to show that conceptual ambiguities underlying current research methods are responsible for many of the research problems. An alternative way of understanding search is proposed with corresponding methods of conceptualization, some allowing mathematical formalization.

1.2.7 Motivations due to the requirement of a new definition of Search

In light of the above motivations it is clear that search requires to re-interpreted to cater for its given function in the sociological sphere as a tool for forming the sociological imagination and in the educational sphere as the primitive function of an universal education. In order to do this the most basic principles of search require to be re-investigated, from its formal definition as a type of physical process which should ideally exhibit scientific interpretation for creation of a 'search science', to its ability to consider cognitive modelling for better understanding of users, to computational modelling of search software, these all require

---

[16] This refers to the way in which a system, observations about the system, and hypotheses are represented and the methods for measuring the truth of hypotheses. In the context of this thesis it should also be taken to mean the relationship between the laws of nature, such as those pertaining to energy, and the modelling methods of QT.

[17] This refers to the way QT employs the concept of states and state changes to represent physical systems and their evolution.

to be addressed in an unified way. This thesis does not provide a conceptualization of search that immediately caters for the socio-educational perspectives but instead it directly addresses the ideas at the conjunction of the scientific, retrieval-specific, and cognitive perspectives, and in doing so it hopes to provide a discursive framework for developing socio-educational perspectives *from within* the formal scientific framework, or be able to integrate such perspectives into the formal computational perspective.

1.3 Thesis Statement

This thesis addresses the nature of the search process and the main problems responsible for IR research being in its current non-ideal state. An outline is given of current work on employing QT to address one of these problems proceeding with an approach to resolve the other causes. In the next section, with reference to the traditional laboratory perspective of IR [IJ05], the nature and scope of the the *evaluation problem* and *user problem* are discussed. Both these problems are dependent on the *definition problem*, which therefore needs to be addressed first. The laboratory view of search is a hindrance to adequate conceptualizations of these problems, thus an alternate view is suggested. The modelling technique in Chapter 5 provide this perspective enabling the visualization of the interaction of different research areas, with the advantage that the definition, evaluation, and user problems can be represented visually and defined formally in terms of *states of search* and the *knowledge of states*.

Chapter 3 concludes with details of all the key problems particularly elaborating the *conceptual problem* of defining search and with that resolved Chapter 4 outlines methods for modelling IR, specifically the changing information need in a search, using the mathematics of QT, further benefits of adopting QT concepts for IR are also presented. Our approach to solving the research problems in IR with inspiration from QT raises several new and interesting questions, suggesting changes in the method of experimentation, and re-defining boundaries between related research areas. In Chapter 5 IR is discussed as a simple communication process. The aim is to use the formalization of the communication process as a (semi-formal) base language in which to express IR in a general form; then to systematically map both the representation and semantics pertaining to retrieval concepts to the appropriate modelling language (such as QT) that is expected to be beneficial for analysis. Conceptualizing IR in the ways presented in this thesis lead to new questions about the search process which are not obvious from the traditional (laboratory) view of search. The questions are of the philosophical, theoretical modelling, experimental and implementation-specific varieties. As the communication process view of IR, once formally defined, resembles several other types of communication, one is led to enquire about what it is about the 'search' process that makes it different from other processes. Is it a certain state in the user or the use of a search program? Or can some other communication between man and machine or more generally an agent and another agent also be classed as search and would it be useful to view it as such? Defining search as being formally different to other communication processes does not seem easy at this point. This raises an interesting but perhaps obvious point: if search is like other processes and hard to distinguish, then can other processes conversely be interpreted as search? In practical terminology, can the process of a user doing a normal activity such as logging onto a machine or typing a document actually be interpreted as search? This line of reasoning can be used to formally reason about incorporating search functionality into different parts of an operating system, something which is already being done in the latest operating systems where the user does not have to "go to search" through interacting with software, instead the software attempts to detect when the user might need to search and facilitates search functionality within their current interaction environment[18].

On a practical level, the perspective of IR provided in this thesis suggests simulation as a natural way to evaluate search systems and can be used to identify if and when user studies may be required, potentially minimizing evaluation costs. Describing parts of the IR process in languages like that of QT suggests implementing IR systems in the same way the corresponding QT systems would be implemented, some of which have been extensively studied but not applied to solving IR problems.

On a research level, the communication model view of IR implies that information seeking activities can also be modelled in the same way, as further agents in the communication. More importantly it suggests administrative changes to IR research methods, so that re-

---

[18] Either within the software being used at the time or in some other convenient way.

search is divided according to the model, an important suggestion of which is to separate user modelling research from 'search scenario' research with the aim of potentially speeding up progress by introducing a parallelism in research. The vision of retrieval research encouraged by this thesis is one in which a researcher formulates a concept pertaining to search, defining it formally using the given formalizations (or derivations from it), then obtains pre-defined user definitions and conducts an evaluation of their concept by simulation. The results of the simulation have a strict formal meaning as well as a practical one and are specific to the simulation parameters. This strictness is in contrast to the ambiguity in evaluation techniques apparent in the laboratory perspective of IR [AvRJ05].

1.4 Thesis Outline

This thesis is composed of seven chapters, the background in Chapter 2, an exposition of a novel conceptualization of search in Chapter 3, an exposition of the use of QT and its inspiration in creating a new view of search (as in Chapter 3) in Chapter 4. A semi-formal state based notation and methodology is given in Chapter 5 which aims to provide a research method for creating new retrieval models, evaluation metrics and strategies, and user models, suggesting how the various perspectives of search could be integrated through use of semantics at the modelling level (as opposed to at the discursive level). Discussion and elaboration of applications of the suggested research method is in Chapter 6, finally I conclude in Chapter 7.

Background (Chapter 2) : This chapter further justifies the motivation, elaborates current approaches for dealing with the query formulation problem (by ostensive retrieval), and briefly discusses evaluation strategies, cognition modelling and the embedding of search functionality in generic software applications. It addresses shortcomings of traditional and current approaches to IR research by introducing the definition, conceptual, evaluation and user problems which are to be addressed by the remainder of the thesis.

On Reduction to the Higher Ostensive (Chapter 3) : In this chapter the conceptual problem introduced in Chapter 2 is addressed by use of a generalization of ostensive retrieval. The principle of higher ostension is introduced to address how a search process should be modelled and the similarities and differences in modelling search from the user and researcher perspectives. In discussing the the concept of the ideal search agent the method of user simulation is suggested to be conceptually and formally necessary in order to address the conceptual problem in IR (and hence all problems as introduced in Chapter 2).

On the use of Quantum Theory for Search (Chapter 4) : This chapter gives some background to the mathematics and interpretations of Quantum Theory. Following from this, various avenues are suggested for conceptually understanding a search process as a physical process. Several modelling methods are presented, firstly techniques for encoding dynamic and static search information in terms of Hilbert space mathematics are introduced. Also included are ways for modelling information need change using group-theoretic methods on matrices, new measures for the utility of a search state corresponding to a combined measure of static and dynamic search information, and a discussion of further techniques for modelling search.

On Search as a Communication Process (Chapter 5) : In this chapter, I present notation that embodies the higher-ostensive principle, and notion of state and state change from QT. The main aim of this chapter is to address the definition problem by this notation such that arbitrary searches can be interpreted and represented in a way that facilitates a QT style understanding, and also a computational (programmatic) understanding.

Applications and Discussion (Chapter 6) : This chapter includes some characterizations of the search problems mentioned in Chapter 2 in terms of this new notation that aids in relating the problems together. In this chapter I investigate applications of the notation introduced in Chapter 5 in modelling structured documents, users, user interactions, evaluation strategies and the representation of the concept of ‘information experience’ that was introduced in Chapter 3. This chapter also offers discussion of the role of the conceptualizations of this thesis with respect to other conceptualizations, relating between concepts in QT, IR and computational concepts.

Conclusions (Chapter 7) : Here I report the main contributions that this work made, and point to some issues for future work that follow from this thesis.

–A philosophical problem has the form: I don't know my way about.

Ludwig Wittgenstein (1889-1951)

2

Background

2.1 Introduction

In this chapter basic concepts from traditional IR are reviewed in contrast to human-human information retrieval (HHIR). References to the *effectiveness of search* are not taken to be purely based on particular evaluation measures such as precision/recall but also on the ability of imitating HHIR which is assumed to be the best form of information retrieval at least in terms of interaction protocol. A search system better exhibiting HHIR behaviour is preferred over one that imitates less[1]. In terms of efficiency the human may not usually be as exhaustive as a search engine but it is assumed that a human augmented with a search system would together form a better search tool. HHIR refers to multitudes of search scenarios, for example from asking for driving directions to discussion of complex scientific topics. An arbitrary HHIR scenario cannot be assumed to be ideal, instead the specific scenario of formal academic research with its protocols and constructs, is assumed to be that in which optimal IR happens. Alternative HHIR contexts are out of the scope of this thesis, but it is assumed that by taking the academic-research context many of the other HHIR contexts are automatically covered due to the comprehensiveness of academic IR. The analysis is further limited by only assuming comparison to an academic-research session that involves one human agent and one system agent (unless otherwise specified) in one search process/session as opposed to multiple users and/or sessions.

This chapter discusses the laboratory model, associated models and their relation to HHIR concepts, as what type of persona in HHIR corresponds to the conditions of representation in the vector space, probabilistic, and logical models? It is natural for humans to think in terms of chance, space and use deduction (and in other terms) corresponding to the probabilistic, vector space and logical models respectively, perhaps at once. Thus the three canonical models by which the system makes search decisions denote but a part of the user's decision mechanism and do so, relative to HHIR, in an artificially divisive way. As a quantum theoretic approach combines these canonical models into one, it is argued that QT approaches for IR result in models that better accommodate cognitive science understandings of human communication in HHIR than the current traditional IR approaches, and are therefore a more suitable medium for imitating HHIR for CIR.

---

[1] How 'better/less' imitation is to be defined is not addressed until later chapters through the concept of 'convergence'.

Initially the three canonical models are presented with a discussion on the concept of relevance, feedback and evaluating IR systems, all from a traditional IR perspective. Ostensive retrieval is discussed in Section 2.3 as an approach that addresses the query-formulation problem which itself is a problem of user-system communication. The user is discussed in Section 2.4 in terms of cognition in order to deduce the nature of the human agents in HHIR and issues related to imitating them for simulation based analysis of search scenarios and for creating more human-like systems for computational information retrieval (CIR).

In Section 2.5 the discussion pertains to the nature of a search system arguing that its definition has been blurred due to search functionality being augmented to different parts of an operating system and manifesting in different ways on web applications; therefore the concept of a system for IR research needs to be re-thought to address these various manifestations in a unified manner.

Finally, the issues raised in analysis of the current perceptions of CIR, the search system, and user, are generalized as problems of representation (syntax) and operational method (semantics) and abstracted in terms of four related problems in Section 2.6 which are to be addressed by this thesis.

2.2 Traditional IR

The traditional model of a search process is depicted by Fig. 2.1 and still applies to most search systems, see [IJ05]. In order to search a data set, the documents in the set must first be indexed according to a document model. The index is the same data set reduced to contain just the information (collection of words, media, and any metadata) about the set required to represent the collection of documents sufficiently according to a document model. Queries expressed by the user are interpreted according to a query model that is implicit in the search request of Fig. 2.1. A matching sub-process follows which takes the query and for each document assigns a value to the association between the interpretation of that query and the interpretation of the document according to their respective models. This association is termed relevance and can be defined in multiple of ways [Miz97]. The results of the matching process are manipulated and shown to the user according to an in-

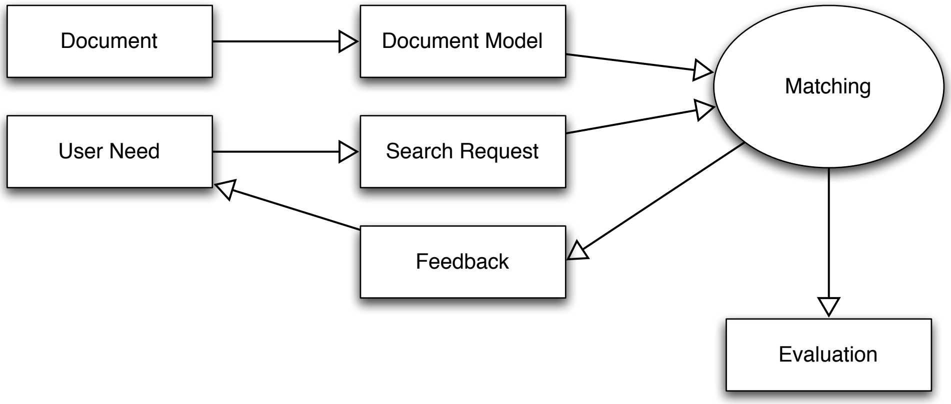


Figure 2.1. Data flow in the Laboratory model

terface model which is outside the scope of the laboratory model. For a simple list interface like that of Google, the results are ordered according to matching values, so that documents deemed most relevant to the query appear higher than the less relevant ones. There are search engines that allow the user to influence the matching sub-process by feedback so that the system's decision about what it thinks the user means by the query, changes; the feedback can be seen as an *implicit query*. The user feedback expressed through the interface is known as relevance feedback and facilitates learning of user interests.

In order to realise our motivation of trying to make CIR more like HHIR two major issues associated with the former are initially addressed. The first aim is to show that at the root of these issues lay a set of assumptions and inadequate modelling that not only limits CIR as it is but restricts its improvement. Secondly to solve this problem it is proposed that the conceptual foundations of CIR needs to be addressed by further investigation of HHIR.

Consider Fig. 2.2 which illustrates the process of the user expressing their information need showing that in order for a CIR process to start there requires to be a transformation of what is initially an abstract concept to a search expression. These transformations represent the process of query formulation which is one of the most challenging activities for the user as addressed in information seeking literature [CPB$^+$96]. This is especially true if this abstract concept, the user's information need (IN), is vague, or if the user has little knowledge about the collection make-up and retrieval environment [SB90]; understanding information need is one of the central retrieval problems. The unprecedented level of expansion of the web has led to a rapid increase in the number of casual users using web

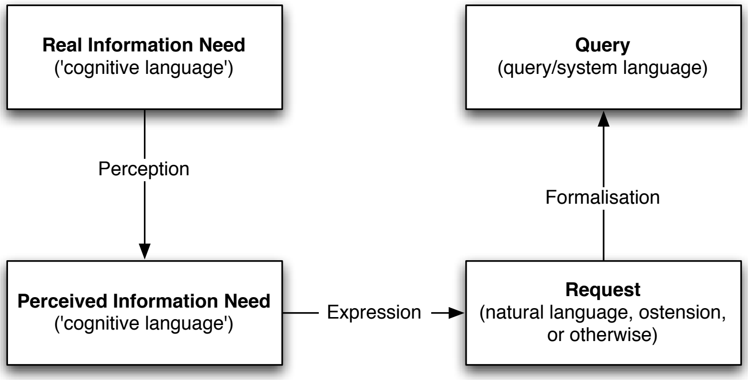


Figure 2.2. The pre-query stages of search

search engines. Most of these searchers are inexperienced and find it difficult to express their IN. Research has showed that nearly two thirds of queries submitted to search sites are only one or two terms [JSS00]. This is especially apparent during the initial retrieval operation, when knowledge of the retrieval domain and awareness of their IN is minimal. At this stage queries are short and imprecise, leading to ambiguity and a shortfall in the number of terms needed to retrieve relevant documents [Har92, San94]. Information need analysis using such evidence is difficult and ineffective. The majority of the users fail to use search systems effectively as they are unaware of its mechanics; they lack understanding of how any output provided by the system is linked to their queries therefore being unable to provide a *good* query.

The Two Key Problems of Traditional IR

In summary, the following two issues can be defined as the main problems with query reformulation and hence CIR:

1. Query Formulation Problem (QFP): Query formulation is not an easy task for users as most often the underlying information need itself is typically vague [tHPvdW96, URJ03].
2. System Understanding Problem (SUP): Furthermore most do not know how the documents are represented and seldom do they understand in a consistent way how

their queries are to be interpreted or how the matching process works[2].

If a user is searching then when can one say that they are no longer learning to use the tool and have started to employ the tool for search? What is the boundary between learning and (effective) use? When a user interacts with search software is it an attempt to understand the tool, express/understand their information need, or is it both[3]? It is difficult to separate these interaction objectives of learning and searching as even the user may not know *if they do not know how* to effectively use a search tool. This issue is yet to be addressed by formal search models and there are also no IR frameworks that allow investigation of such notions in a formal way[4].

On using a web search systems such as Google, the users suspect that there is word matching at play but often the results are not the expected ones making the nature of the experience explorative. A user's need may change dramatically in these situations due to the exposure to new information. This often leads to a reformulation of the initial query to either make it more precise now that they know more about the collection make-up, or to make it reflect the interest in newly seen documents, or a combination of both. Inherent in such exploration is the user forming an idea about how the retrieval system answers its queries. For the average user, such an idea is at best slightly more advanced than the suspected word matching. Compare this to the HHIR situation of conversing over email which allows each party to develop far more complex ideas about one another[5] even with as little data exchanged as in the human-computer situation. It is due to intuition and multiple communications that each party develops an idea about the others behaviour/nature which they expect to be somewhat consistent. However, in the human-computer case, although the system understands the user in a limited way the user does not have enough support/evidence to form a consistent 'opinion' about the system.

---

[2] The user's difficulty in formulating queries as denoted by the QFP is not only a IN transformation problem but also due to lack of knowledge of how to use a system effectively, hence QFP and SUP are related.

[3] Compare this interaction between user and system with a discussion between two technical discussions between academics in HHIR where IN is assumed to be well expressed.

[4] QFP/SUP issues are addressed by the notion of convergence in Chapter 5 and especially Section 5.4.8.

[5] Thus in HHIR the equivalent to SUP is the human understanding problem. It is addressed by models of communication. The process of sending email also includes the equivalent of the QFP that refers to natural language expression.

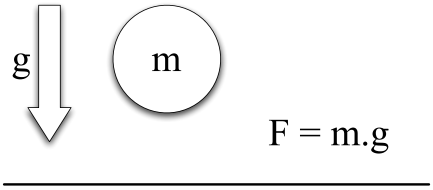


Figure 2.3. Simple physical model with underlying theory

Relevance Feedback Relevance feedback (RF), the idea of extracting more information about the users need by asking more questions or otherwise, is one way to investigate QFP. It is a way to address the HHIR analog of QFP whereas SUP is analogously associated with both personality and presentation. The idea is to allow the user to learn how the search tool works and how to express their need through continuous analysis of their information need and if feedbacks of other users are also expressed (through recommendation subsystems) then to teach users to learn expression according to how other people appear to express that same need[6]. Section 2.2.1 details relevance feedback, and Section 2.3 discusses how OIR approaches these problems.

Relationship between components of a search Unlike relationships between components in a physical system, for example as described by equations denoting physical laws such as in Fig. 2.3 which are made apparent through a prior theoretical framework, IR research does not exhibit general frameworks to deduce such relationships between its components. In the physical system the effect of modifying one parameter on other parameters can be predicted, in an IR evaluation the user is a parameter but there are no ways to determine the effect of its modification on the evaluation. Since there are no formal methods to relate users, the potential effect on user judgements of modifying a component, is generally too unpredictable relative to the physical case, prior to experimentation.

As concepts of effectiveness of a search system are inevitably tied to the characteristics of particular users employing the system, an approach for creating theoretical apparatus for IR with the ability to formally compare components would need to tackle the user issue. If instead there were a way to work with formally specified abstract[7] users which approximate

---

[6] Which would be deduced through HHIR between people observing other people using search.

[7] Abstract in that they are automatons which may not represent a real user or if it does then it is inevitably a limited model of a human user.

real users then experiments can be duplicated and verified. The problem would then be one related to the effectiveness of the approximation of abstract user models. With an accepted user model, the abstract user would then be a controllable experimental factor. However, there are no general formal methods for abstracting user behaviours for creation of abstract user specifications, and it is not common practice in IR research. Instead the research literature uses brief and informal natural language descriptions for users, i.e. 'the users are university students with moderate experience in searching'. A practical advantage of formal specification is that it would allow relatively economical user simulation type experiments and provide definitive results for a specific user specification (see [WRJvR05]) which can then be verified. With formal specification the effectiveness of a search engine can be identified with specific user types and evaluation results can be reasoned about in terms of the user specifications.

2.2.1 Relevance and Relevance Feedback

One could suggest that with the QFP it would be useful if the user were aided in some way to make describing his information need easier. Relevance feedback techniques address this problem so that short, ambiguous queries that are evident especially at the start of a search session can be augmented with further indications of the IN. This typically ineffective initial query suggests that in RF systems this query should not be regarded as a good description of IN. During this initial 'experimental' phase of search [SB90] the user selects documents that are relevant, and terms from these documents are used to automatically devise a new query, which is then resubmitted to the system.

The whole RF process is a controlled, automatic means of query reformulation performed instinctively by the searcher when dissatisfied with the intermediate results. Its importance can be realised to a greater extent in the image domain where it directly addresses the semantic gap problem [URJ03]. RF is quite a *natural* process in that it is what happens in human communication/HHIR[8]. Judging the effectiveness of an RF approach requires us to specify an evaluation technique and a retrieval model. Using the traditional evaluation method would allow a result set containing all of the relevant (and none of the irrelevant)

---

[8] RF in the HHIR context denotes clarification of spoken concepts through further conversation.

documents in the to be used as a pre-judged set.

Initially, research into RF was done on the basis of two IR models; Vector Space and Probabilistic. The vector space model was first used by [Roc71] where RF is represented by a query vector changing per RF coming *closer* to the vectors representing the relevant set, as measured by a distance metric. The assumption here is that the more similar some document is to some pre-judged 'relevant set' the more likely it is to be relevant [RLvR01] with respect to IN. The use of the probabilistic model [RJ76] for RF, and in particular term weighting, has been more widely addressed. The model ranks documents based on term distribution information of relevant and non-relevant documents. Early work [Jon79] showed that probabilistic weighting schemes were useful for relevance feedback, especially for term weighting. Extensions detailed techniques for incorporating new terms rather than just reweighing [Har80, SvR83] but their successes are dependent on the corpuses.

Users perform RF by interaction that can be explicit where the user is asked to judge the relevance or non-relevance [CBK96, BCC+98] of some document or select terms for the Query Expansion (QE) process that implements the feedback process [SB90, Rob90]. This shares with query formulation the disadvantage of inflicting a (cognitive) burden on the user since it is difficult for most users to numerically assess the degree of relevance of one document, which presumes considerable knowledge of the retrieval environment. A less-distracting possibility is through implicit feedback, i.e. observing user interaction with the system [WRJ02].

Nearly all examples that use these feedback approaches assume that a user's information need is static and provide no way to update the system's view of user's needs in accordance with changing IN. This greatly simplified view of reality does not consider the change that occurs in user's search *goals* corresponding to changing information needs which in turn are reflected by their *actions*. An exception is the implicit feedback approach in [WJ03] which interprets IN change between steps by analysing the statistical rank correlation between query terms at those times. However, there is no justification for interpreting the changing ranks of query terms as IN change, it is an assumption. Furthermore no approaches consider that the dynamics of IN change also varies. For instance, an abrupt or *high momentum* change can be triggered as a result of having come across something interesting that the user has not even considered at the beginning of the search while *low momentum* changes refer to less drastic change of need/action. The Ostensive Model, which steers away from these assumptions aims to capture the intentionality of an information need that is assumed to be developing during the search session [Cam00]. The relevance judgements in this model are obtained implicitly by interpreting a user's selection of a particular document as an indication that this document is more relevant than other selectable ones. As explained in Section 2.3, this indication is used to learn and adapt the query which then consists of a combination of the features in selected documents.

2.2.2 Evaluation

A search system is evaluated according to its effectiveness and computational efficiency. The effectiveness of a search system is usually deduced on analysis of two features, its ability to correctly associate query and document (to judge relevance), and to adequately present results on the interface. Effectiveness of a set of relevance judgments is traditionally deduced using precision/recalls measures[9] to compare against prior human judgments on the same queries and in the same document collection, the majority of evaluation experiments are performed this way. The method of evaluating result presentation, or interface models are equivalent to human-computer interaction evaluations typically employing questionnaires and usage log analysis. IR research tends to focus on improving the effectiveness of relevance judgments by the creation of novel models for individual components or methods for combining different models. To test a new measure for matching a query to documents, one would select a test collection of previously collected user judgments and run precision/recall experiments for queries in the collection to deduce effectiveness. Such user studies take justification from social sciences and psychology and the results provide more of an empirical justification than a formal one.

---

[9] Precision is $(\frac{RETREL}{RET})$ whereas recall is $(\frac{RETREL}{REL})$ where RET is the number of documents retrieved and RELRET is the number of previously judged relevant documents found in RET.

2.3 Ostensive Retrieval

The Ostensive Model (OM) of developing information needs [Cam00, CvR96, URJ03] adopts a query-less interface, it combines the two complementary approaches to information seeking: the query-based and browse-based search method in which the user indication of relevance of an object-through pointing at an object is interpreted as evidence for it being relevant to his current information need. This method of capturing relevance judgements is supported by Quine in [Qui60, Qui69] which explains that to capture a cognitive concept (such as IN) one needs evidence through observations[10]. These observables (observable ostension) should be defined to be non-composite simple actions and the interaction objects defined such that the motivation of the actions has an interpretation which is as straightforward as is assumed is possible. Finally, that a set of elements of the language must be compatible, in that it should make sense to associate/group them for the purpose of generalization or collective interpretation of pertaining information need. Quine suggests that effective ostensive expression corresponds to a simple 'ostensive language' so that the effect of the SUP is minimal in the context of learning the interface. In this case the QFP can be approached by a user who has a simple tool so that they know that repetitive simple interactions ought to lead to IN satisfaction since they can plan with it and around it. With more complex interfaces, as with complex tools for physical tasks a typical user's planning of use of the interface increases and/thus their use becomes difficult.

The OM therefore allows direct searching without the need of formally describing the IN. The ostensive path that results from a search process also visually represents the dynamic nature of information needs. Through search the user is exposed to new objects changing their context and developing their knowledge state[11].

In treating a recently selected document as more indicative to the current information need than a previously selected one, the OM recognises that the relevance of documents/needs change over time[12]: "The Ostensive Relevance of an information object is the degree to which evidence from the object is representative/indicative of the current information

---

[10] The next chapter addresses the philosophical significance of ostension.

[11] The graph on the ostensive interface Fig. 2.4 is a visual representation for 'dynamically changing need', it is also a useful visual metaphor for state change as elaborated in the next chapter.

[12] Change over time refers to change over the time period of a search session.

need" [Cam00][13]. In the ostensive browser [Cam00, CvR96] a user starts with one example document as the query, and as a result a new set of candidate documents (top ranking documents according to the similarity measure used) is presented to the user. As a next step, the user updates their query through selecting one of the returned documents, which now consists of the original document and the selected document of the set of returned candidates. After a couple of iterations, the query is based on a path of documents.

The ostensive browser is similar to the Path Model described in [CRB98] for activity-centred information access, its emphasis is on the user activity and the context, rather than the pre-defined representation of the data. A path represents the user's motion through information, 'a spatio-temporal map' which in the OM is used to build up a representation of the *instantaneous*[14] information need. Thus the visual model makes apparent the users' activity to themselves; the visual model is a cognitive stimuli that then encourages a particular type of user interaction[15].

As the whole path is kept visible, users are allowed to jump back if their current path is not judged to be fulfilling their needs. From there a new path can be explored, resulting generally in a tree-like structure, technically a graph since backward link are allowed. Linking to a former document would occur if by the similarity measure the IN at a later time is well associated with that of the earlier time. The OM also incorporates uncertainty into its interpretations of user actions modelling the idea that the user is uncertain of their own needs and even when aware have difficulties expressing it, thus their interactions are inaccurate depictions of IN.

Effectiveness evaluation of the ostensive approach has been undertaken only for the image domain [URJ03]. The OM uses simple continuous ageing functions such as 2^{-t} with which to scale the relevances of selected documents over time and hence decrease the contribution

---

[13] As the tree has increases in length it acts as a visual metaphor of ageing, such that one may propose that the documents to the left are less important assuming IN changed; this is of course very presumptive.

[14] Note also the 'instantaneousness' of this, compare to instantaneous speed versus the expression of speed in terms of derivatives in calculus, thus it can be said that the ostensive interface captures IN in a way conceptually similar to the way tangents are used on mathematical curves to deduce their gradient. As addressed in Chapter 4, this instantaneous observation style of the ostensive interface naturally corresponds to the measurement of properties of a search state using QT methods.

[15] As the user is shown their past choices they are given a map of their IN change from the beginning of the search, this is assumed to encourage the user to think and behave (interact) differently than if they were given the usual list interface for example.

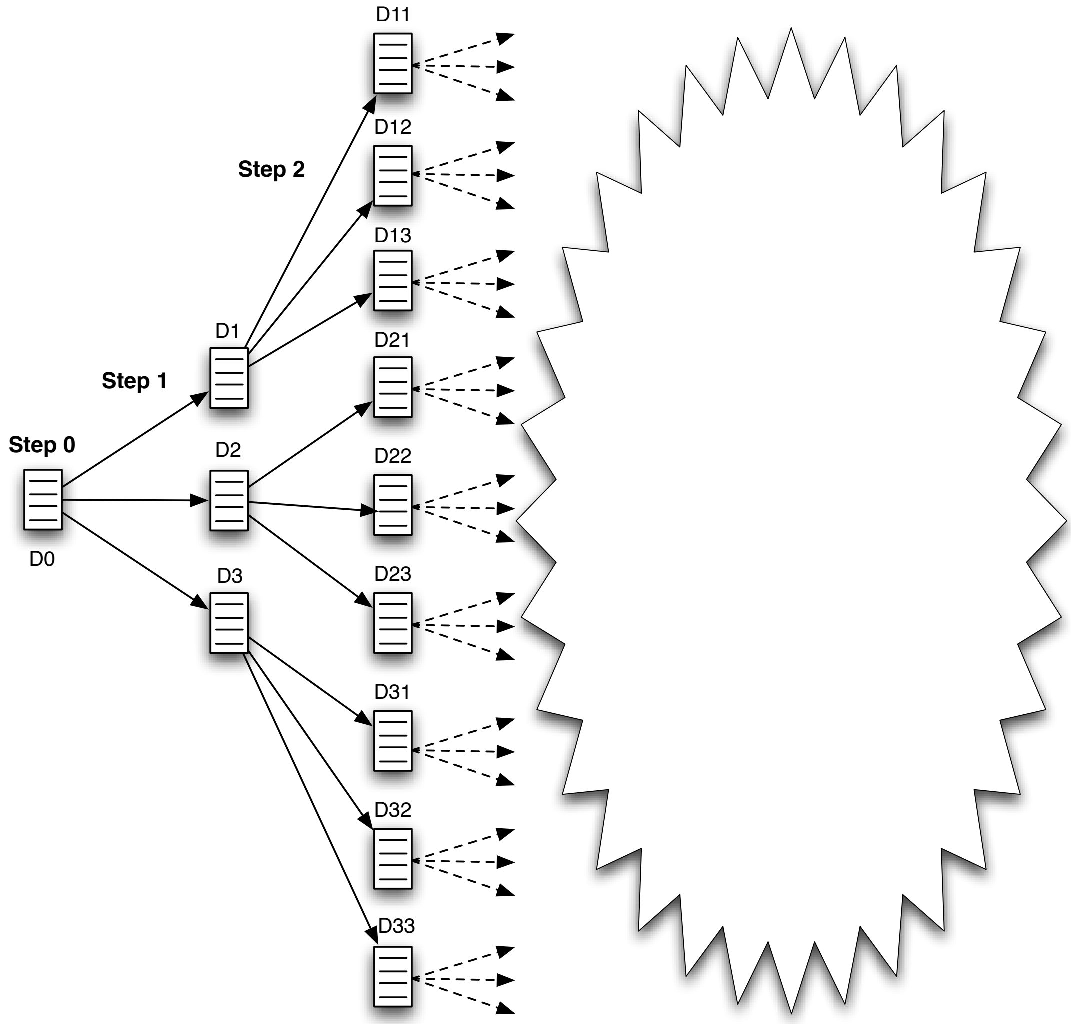


Figure 2.4. The ostensive path

their features had in the implicit query reformulation. It is one consistent way in which the model interprets the implicit feedback, a way which maybe understood by the users in the same way as they interpret Google as word matching alleviating to some extent SUP. The intuition of older search behaviour being less relevant in this way to the current situation is however a similarly simplifying assumption as it means all documents decrease in relevance at the same *speed*, for which there is no justification. Both the concepts of matching words and treating older evidence as less relevant (in the above way) could be character traits of two participants in communication for the purposes of retrieval (HHIR) but are too primitive to describe communication between human agents. Thus ageing is too presumptive, in reality a relevance reduction cannot be same ‘speed’ for all topics. Ideally there would be a comprehensive set of rules and a reasoning system for interpreting the communication which would change relevance weights over time according to supposed IN and user interaction patterns.

In summary, the OM presents a retrieval system that alleviates the word based QFP and provides an ad-hoc heuristic that is too presumptive to address SUP in a reasonable way. Instead a consistent set of rules are required that are intuitive in that the user can personify the system with them and also assume observably consistent application of these rules in the search process. If the users can understand how the retrieval system operates, corresponding to understanding its 'character', they could use their (relatively) superior reasoning to make use of it efficiently as a tool. Effectiveness in such a context depends not only on the consistency of these rules but also on the consistency of the association between them and the interpretations of actions. This can be said to be analogous to consistency of personality in the human-human case. The consistency of this association in turn relies on a good understanding of the semantics of each action for which one must ideally have a good model for cognition. For the purpose of imitating HHIR one should ideally personify the system, designing it with consistency of personality meaning the way the system behaves with a user, so that the user spends less effort in dealing with the 'artificial' character of the system. In the proceeding section these cognitive aspects are addressed, as if HHIR is the goal then its particulars require to be elaborated. This is especially the case for those particulars pertaining to information need which in turn is related to the concept of relevance.

2.4 Cognition

2.4.1 Introduction

The preceding section indicates that CIR is currently an inadequate imitation of HHIR. The agents in HHIR are complex cognitive agents who *know* well about each other. In the language of CIR [Rob03] each agent in HHIR has a *good user model* of the other. Each agent also stores their 'ideas' in some way reasoning between them for the *purpose* of expression or internal derivation ('thinking' or 'internal flux' according to the OM [Cam00]). The diagrams in Fig. 2.5 highlight the relative characteristics of agents in traditional search

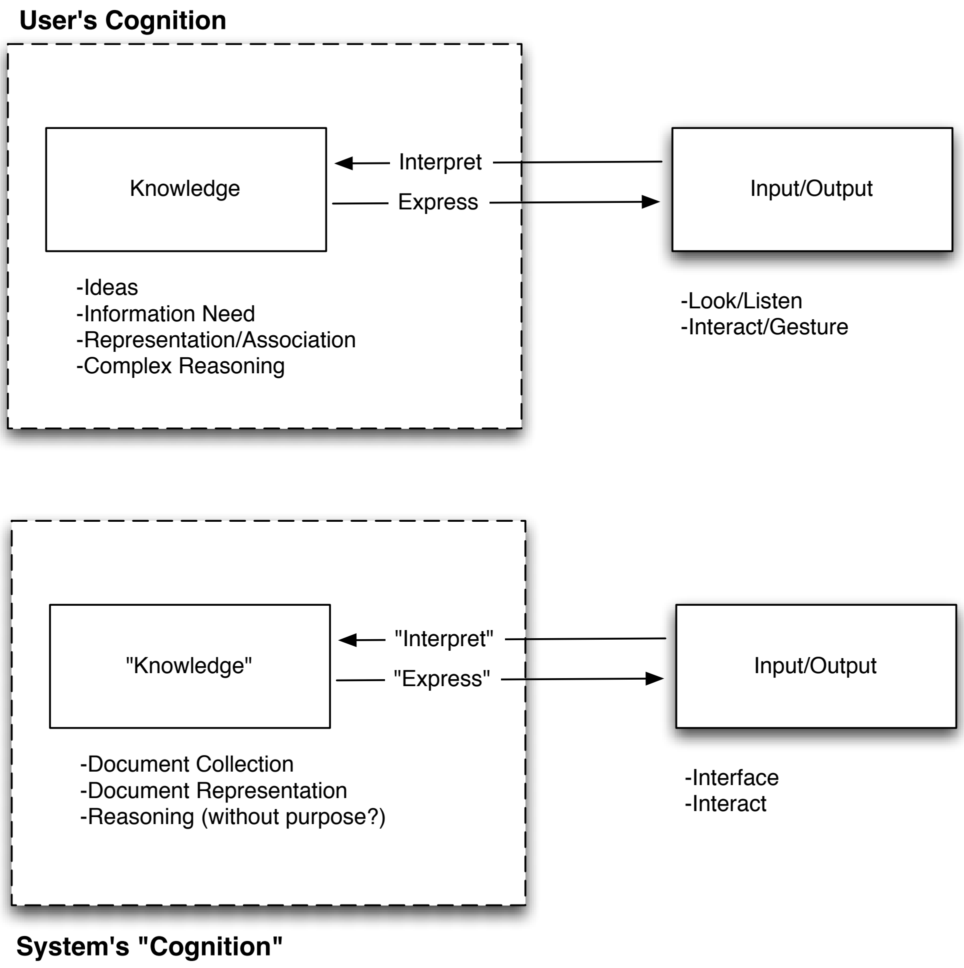


Figure 2.5. System imitates human

in terms of the information held by each agent and reasons for interaction. As will become clear, purpose is an important issue for the system in our framework. It is something that needs acute definition and is related to retrieval strategy, it is also what limits similarity between man and machine in the CIR and HHIR context. The aim of this section is to discuss the characteristics of agents involved in HHIR that CIR aims to imitate. A non-technical analysis of cognition in the context of retrieval is presented followed by a set of requirements for imitation.

During a search process interaction occurs followed by reasoning which interprets the input affecting the knowledge in some way (Fig. 2.6), as supported by OM. In the HHIR case the expression/interaction languages, including natural language and gestures must be mutually understood for the *purpose* of answering questions. In meeting this singular purpose such communication is always at the level of the agent who is less learned in the language of

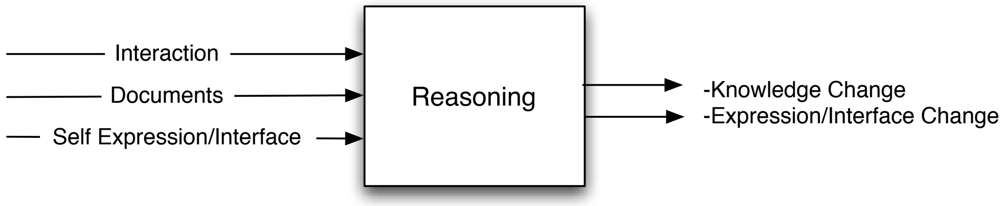


Figure 2.6. Reasoning Process

interaction and expression; consider the analogy of a local giving directions to a foreigner[16]. For retrieval the machine is assumed to be the weaker agent thus contributing to SUP and the loss of accuracy in information need expression thereby reflecting QFP.

An attempt to improve 'dialogue' between agents in CIR to imitate that of HHIR requires an interaction/expression language and a way to judge its effectiveness. As indicated in Fig. 2.5 and Fig. 2.6 the reasoning process and 'knowledge center' of cognition are closely tied to input/output, thus instead of investigating optimal interaction/expression languages directly the focus is on modelling these cognitive elements with the intuition that this will lead to a better understanding of language.

2.4.2 Structure of Cognition

In human cognition knowledge in a generic sense consists of concepts which are *associated* or *correlated* with one another in a way more general than documents in IR document models, this association structure is then used to reason. The first diagram of Fig. 2.7 shows symbolically that the reasoning between the association is the 'On A' relation. If only the proposition 'Person is on a bus' was given then clearly an association is there between person and bus. Any such association generates a statement even if it is the trivial association 'A is associated with B', while every statement associates (not necessarily distinct) concepts.

Associations also denote correlation and have a concept of strength as indicated by the distance in the second and third diagrams of Fig. 2.7 that implies closeness of the friendship relation. The strength or degree of association is a property of the association but can

---

[16] Where the foreigner does not know the local language very well.

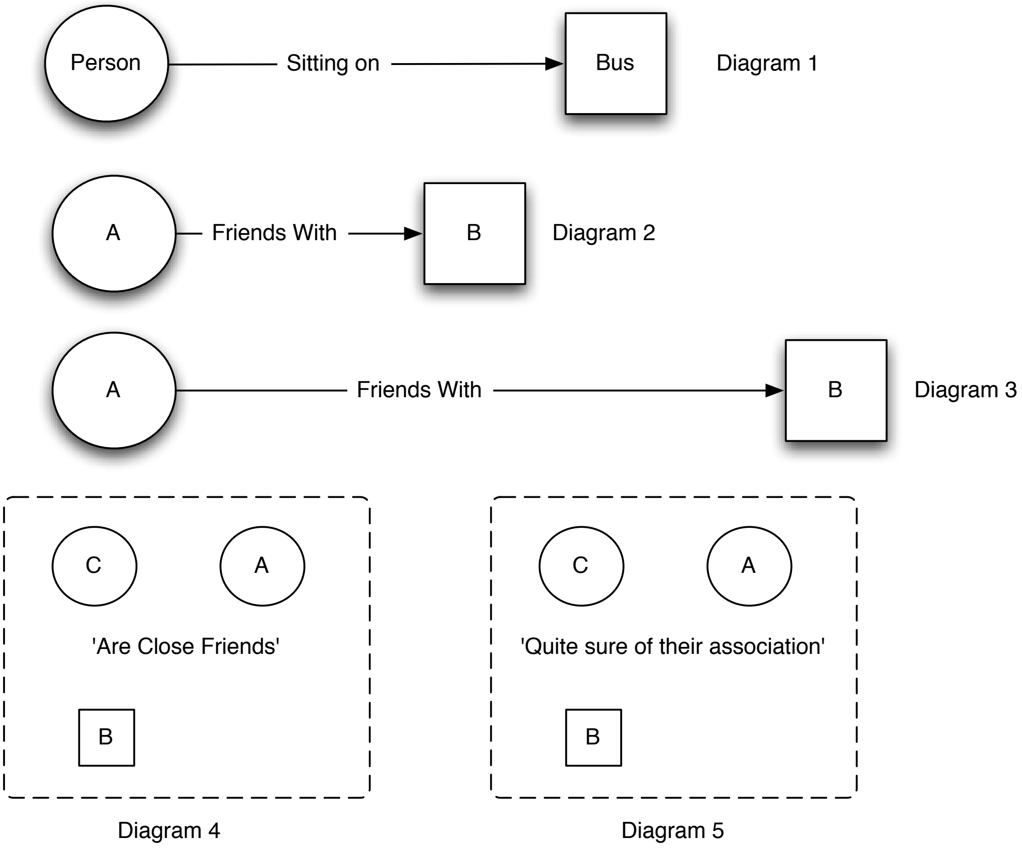


Figure 2.7. Association of Concepts

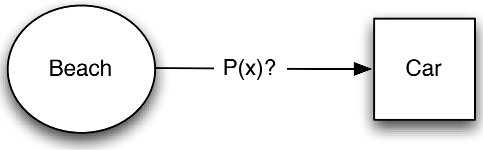


Figure 2.8. Uncertain logic from associations

also be used to generate associations as depicted in the fourth diagram of Fig. 2.7. Finally there is also the notion of uncertainty attributed to associations which is different from strength as it is an indication of certainty one has about the reality of the association and its assigned strength. Uncertainty is a property of an association and also a concept that *generates* an association as shown in the fifth diagram of Fig. 2.7, it need not be a binary property and could entail a numerical value.

Associations, their degree, their uncertainty and the logic/reasoning that results (or vice versa) are simply technical characteristics of concepts as they naturally exist in cognition.

Take for example a low level HHIR process that aims to imitate CIR where one agent observes another looking at advertisements in a tourist office. If the 'question answerer' in this case spots the other agent looking at two topics, allowing $P(x)$, a logical statement to be formed about the association of these topics (Fig. 2.8). This must be uncertain at some point, and would therefore be assigned a (not necessarily binary) weight denoting the potential of the truth of $P(x)$. In CIR, say the ostensive interface of [Cam00], interests are similarly detected and the document model is then asked to retrieve information based on the 'reasoning' inherent in the binary voting model coupled with the ageing normalisation. If HHIR is to be imitated then initially documents would be represented in a way similar to how concepts are represented in cognition.

2.4.3 States and State Change

The associations in cognition and their properties change over time. This could be triggered by input, mentioned in OM's, as well as 'internal flux' [Cam00] but distinguishing change by either label is difficult as it's usually a combination of both. One way to study changes in cognition is to look at its *state* at time intervals. Each state is a *label* denoting all there is to know about cognition at a certain time. One can then look at how each state changes by considering in turn the associations and properties of associations that change over time. A state change can be measured by the proportion of associations it modified and a quantification of the total degree of this modification. Change can also create new associations, for example new concepts can now be associated with existing ones, any measure of change must also take this into account. Information need is a subset of all cognitive associations. During a retrieval session IN change is the growing/shrinking of this set as well as the internal adjustment traditionally implemented in CIR as term re-weighting and query expansion respectively.

CIR traditionally simplifies its modelling of such a general change of associations, association strengths and uncertainties, into three distinct frameworks. Associations and their strengths are modelled by vector spaces and the metrics within while uncertainties by classical probability. Uncertainty is commonly related to *relevance*, a notion used to justify expression on the part of the retrieval system. Neither of these models support theoretical derivation of logical statements for which the logical model of [vR00] is required. Similarly [vR00] does not support general 'association' in the same way as the vector space model, instead it is an implicit notion based on the logical uncertainty principle. This state change is further discussed in [AA94] where the relation between stimulus (input) and changed state is addressed under varying contexts. Further work in [AG02] makes the conclusion that *representational* theories of concepts, such as that by [BOB82b, BOB82a], are especially good at modelling *causation*. However, due to the nature of these theories they are not good at modelling *cognitive state changes* or for addressing the way concepts within cognition (such as information need) change over a period of time. In [AG02] it is explained that such theories are also not good at modelling and predicting state change *semantics* under "atypical contexts or correlations that result from natural creative blending of concepts". In the latter quote 'correlation' denotes an unexpected or non-obvious link association with respect to what classical probability or logic would predict. Therefore the use of classical probability to model uncertainty is also questioned and it is proven that *context induces non-deterministic change of cognitive state introducing a non-classical probability on the state space* [AA94].

For the association/correlation relation of *conjunction* it is found that conjunction among concepts, during a state change, can create new concepts with properties not present in its constituents[17]. This cognitive behaviour is not accounted for by classical probability approaches which require conjunction to mean a *closed* cartesian product between state spaces representing concepts [Rob03]. Upon stimulus/input a state changes and is then known as a *state of potentiality* that is representable as a non-classical disjunction of states or concepts (that refer to potential next states). In summary, to better resemble HHIR one must be first able to represent concepts, a set of which define information need. Secondly one must also be able to model the change of cognitive concepts/IN. Currently no single CIR framework resolves these issues that are central and pre-requisite to imitating HHIR.

---

[17] This refers to concepts which exhibit properties that are non-trivially distinct from a mathematical combination of their constituents. The 'guppy effect', where the concept of guppy does not exhibit a strong association to pet, nor to fish, yet it is known to be strongly associated with the concept of a 'pet fish' [OS81].

2.4.4 Purpose of Cognitive Agents

Before proceeding to a framework that takes into account the cognitive phenomena covered in the previous section an outstanding limitation is mentioned in this section. In human-human information retrieval the purpose of each agent is not strictly restricted to fulfilling the others information need. In trying to imitate HHIR it requires to be assumed that the particular human case to be imitated is that where an agent's purpose is only purpose is to provide services to other agent(s). For example, the case where the *server* is a call center employee determined to answer the callers questions for the sake of doing so, and not for wages; hence one of the agents only exists to serve the other. Every association, reasoning proposition, uncertainty value attributed to the cognition of the server should aim to fulfil this purpose. The complexity arises when one considers that an agent will have to address multiple retrieval sessions and users. Does the purpose then become to serve each exclusively, that is to forget the state of each session upon completion? If so then this could be contrary to optimally serving the questioner in some search process as keeping the cognitive state between search processes[18] could improve the search's effectiveness. It is not immediately clear if learning from one search would help another for some set of search processes, this *high level* retrieval strategy would require to be investigated for samples of search processes in order to deduce its effectiveness. Another difficulty with turn-based search strategies[19] is that of implementing *internal flux* [Cam00] for the system agent which can also be viewed as a high-level retrieval strategy. The idea of a system changing state outside a search process[20] would again require the purpose of the system to be called into question. A similar strategy in CIR is incremental latent semantic indexing between searches to take account of user feedback in forming new sets of semantics. The purpose of this strategy is to gather understandings of the information needs of prior users and use them to aid future searches. It is a form of collaborative filtering by collaborative deduction of latent semantics in data. Thus internal flux and incremental updates of semantic interpretations of data, are examples of keeping and modifying state in between

---

[18] This refers to collaborative filtering where user recommendations, or more generally user interactions are kept over time and used for data classification.

[19] Where the state changes occur in sequence first in one agent then in the agent that is interacted with, as opposed to states changing in agents in parallel fashion.

[20] Meaning outside an explicit search process. The next chapter re-defines search to include processes that traditional IR would not consider search.

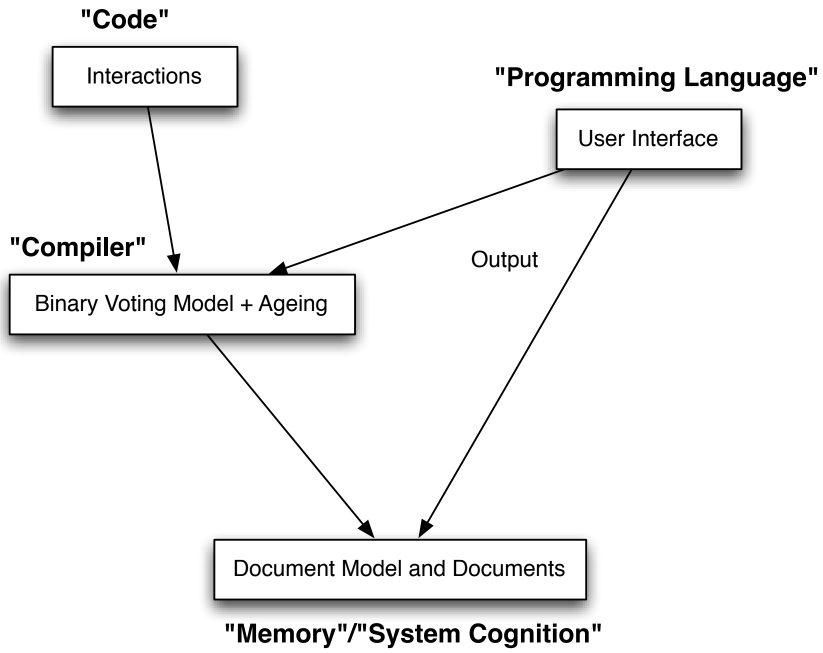


Figure 2.9. The OM as a cognitive computation device

search processes.

2.4.5 Ostensive Model as Cognitive Agent

It is assumed at this stage that the ostensive model is a generalisation of many other interactive search models or can be easily modified to be so, this is further elaborated in the next chapter. In the proceeding the ostensive agent is interpreted as if it were a human agent, but it is initially easier to consider it as a general 'cognitive computation device'. In an ostensive search process the user interaction is essentially 'code' written in the interaction language provided by the OM which are the paths of Fig. 2.9. The user interaction is simply by mouse click indicating relevance.

As the search process continues this *ostensive agent* changes its set of associations. If a document on cars was clicked preceding a selection of one on houses the strength of the association between these documents is said to increase. In fact the association weights between all documents that are constituted of terms common to these are changed. Uncertainty values are not differentiated to association strengths according to the underlying probabilistic model. One can therefore say that the uncertainty is also affected as in human

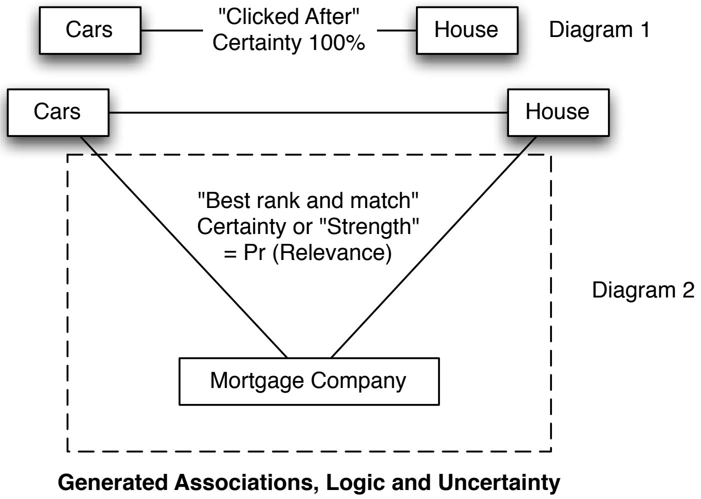


Figure 2.10. Types of State Change in OM

cognition.

During a state change the trivial association of Fig. 2.10 (Diagram 1) is formed along with the reasoning behind the association, a logical proposition, analogous to Fig. 2.8. This association is assumed in the generation of further associations all implicitly happening in the 'mind' of the ostensive agent. Some of these cognitive associations are then expressed or *recommended* (diagram two of Fig. 2.10) as elements of the path on the interface.

Path based expression of the ostensive agent is very close to a representation of the agent's 'cognition', if the cognition is viewed as in the previous sections. As the underlying model is probabilistic it does not allow us to distinguish all the properties directly related to those properties of concepts in human cognition. At this stage paths are taken to be the baseline symbolic and semantic representation of the ostensive agents cognition. It is assumed that there is little or no information loss in translating from the probabilistic model to paths. The 'information' referred to here is that relevant for describing the properties of cognition that can be extracted from the underlying model. The strength of this assumption is questionable at this point but it allows an initial theoretical setup for modelling retrieval with respect to cognition. Its importance overall is minimal as our framework is not tied to a specific underlying model the provides theoretical control over the nature of the information loss between the interface and model/ostensive agent's cognition. The proceeding section discusses further theoretical aspects that facilitate a formal approach to the problem.

2.4.6 Quantum Theory for Cognition

Following from the discussion of states and state change in Section 2.4.3, quantum theory (QT) is suggested as a formal technique for modelling cognitive phenomena with influence from [AG02]. Quantum theory is a way to model (physically) quantum phenomena. Using the QT formalism for cognition corresponds to using modelling techniques taken from the micro-level and applied to the macro or abstract levels in which concepts become analogous to quantum systems or particles, or in the general sense, they become entities with properties that interact with one another and change over time. This formalism can be used for cognition as it accounts for a more general family of cognitive phenomena than any other model to date, including those elaborated in Section 2.4.3. It is also one of the most rigorously studied scientific theories over the past half century thus one can be confident about its internal consistencies. The basic mathematical formalism of QT consists of a complex numbered vector space with an inner product known as a Hilbert space, augmented by algebraic structures that are characteristic of the space and its subspaces such that logical reasoning can be defined between propositions that are represented as subspaces. The inner product measure on the Hilbert space is also used to derive a probability or sets of subspaces representing outcomes (or resultant/implied propositions) of an event (or initial proposition) also represented as a subspace. QT employs the Hilbert space in a way that links the inherent geometric measure of the space as represented by its inner product function with a probability, the subspaces with propositions, events and physical subsystems, and then associates uncertainty of logical propositions (subspace) with inner products between subspaces. Vectors, or more generally, subspaces, are made to correspond to states of quantum systems or cognitive states [AG02]. The probabilities can be used to convey the uncertainty that some state (subspace) will change to another. Studies of representing natural language processing (NLP) in the Hilbert space formalism of QT [WP03, Wid03b, Wid03a, Che02] are additionally relevant for IR as it directly pertains to document and IN modelling. The analysis in [Che02] suggests that the absence of purpose/intention (see Section 2.4.4), an important element of cognition, limits any agent attempting to comprehend language:

> “.. All these difficulties, it seems to me, have to be traced back to the absence

> of an adequate account for intention in both the classical symbolic and the classical physicalist theories of language. The implication due to intention can be profound. For example, we come to an anticipated application of quantum mechanical NLP/AI in automatic agents, to which the intention of the host (a human user in this case) is to be understood. Without intention, it is hardly possible to talk about understanding. This is a crucial difference between a handy tool and a competent agent..." p. 182

[Che02] further supports our justification for investigating HHIR:

> "..In fact, for us as humans, a state of affairs is seldom a fixed and mechanistic representation. It is rather a dynamic and living whole that makes sense, most of the time, only to us. Evidently, the most suitable implementation of states of affairs of an agent are those that are genuinely similar to that of a human." p. 183

2.4.7 Summary

The necessity of using the QT formalism for IR was addressed initially in [vR96] and recently discussed in depth in [vR04]. The appeal of QT as detailed in these works are initially the mathematical tools that comes with the formalism. In order to imitate HHIR for CIR there requires to be a framework which allows modelling of concepts, and associations between concepts, that allows specification of proximity or semantic distance and uncertainty in associations, from which logical statements to be generated about concepts pertaining to a user's information need. Recalling the complex nature of information need [BOB82a], and especially that of changing information need as discussed in Section 2.3 the ideal framework would allow representation of IN so that one can deduce the effectiveness of retrieval models with respect to an IN and design models to accomodate particular INs or IN types. The QT formalism provides an initial structure embedded in a Hilbert space that can be used to represent concepts, their proximity, uncertainty between associations and logical assertions. It also allows modelling of traditional retrieval concepts as [vR04] shows, however, whether the varying concepts from CIR, HHIR and cognition can be adequately represented in a QT framework from which new models and insights can be generated remains to be examined (addressed in later chapters).

2.5 Defining the Search System

Traditionally, the search system is software that is explicit about its role as a search engine. It contains a matching subsystem, a corpus that contains data from where data is to be retrieved, an interface that accepts queries and to which results are sent for presentation from the matching subsystem. As on the web the corpus can be continually changing requiring the document representation which in the practical sense is the index, to be updated over time. Data to be retrieved is of increasingly varying types, first simple documents or web pages to multimedia, applications, and specific types of web applications such as social networking applications[21]. For example, on social networking sites people search for applications which contain no information themselves in the traditional sense but are a middle step for the user to find particular types of information about contacts on their social networks. Once one of these web applications are retrieved they are used to extract information that is perhaps not available to general search engines as the information is only available over a particular communication protocol embedded in the applications. The HHIR equivalent of such a scenario would be a searcher being lead to a person who can provide them information only available from this type of person, for example, this could refer to knowledge only available to specialist academics. Extending the case of retrieving web applications on social networking sites, consider the computer usage scenario of looking through the filing systems of one's computer on an operating system. Is this computer use a type of search itself where the implicit query consists of interactions due to normal computer use, with the operating system playing the role of an *implicit search engine*? Recording general use of particular applications for implicit feedback is useful for predicting the user's information needs [DCC+03] but would recording general user behaviour when using a computer be of use in deducing their information needs when they explicitly want to search? Thus can one say that the user looking through a menu in an application

---

[21] Almost as if one is searching for a particular way to connect to people known to them, to engage in a particular communication protocol.

denotes implicit ‘searching’? These questions partially denote what will be known as the **conceptual problem** of IR which pertains to the definition of search which is the subject of Chapter 3.

As the web becomes structured around higher level semantics, going from a website to a service or application, it seems as though the set of applications on the web are collectively approaching a distributed operating system. Information, whether it is in a document, parameter in a function call, is the basic constituent of a world-wide operating system (WWOS). The search page in the browser or the browser application is the centre for navigation[22] where distances in the context of physical location, ‘ownership’, and specific functionality are blurred and replaced by semantic distance. The switching between programs on the ‘internet OS’ is through hyper-links, and bookmarks, and its ‘central space’ is often a blog or a search engine site which acts as a ‘menu bar’ in ranked list form listing the ‘programs’ (sites/applications) that maybe used (visited). In contrast, traditional operating systems are moving towards a web structure as there is a blurring of physical locations of objects, associations between objects be it user documents or applications, are increasingly based on semantic relationships between objects in order to simplify data organisation for the user. It is as if there was a trend in removing the semantic gap, or in terms of information need transformation, it seems that the eventual query language which for general computer use denote arbitrary interactions, is heading towards the cognitive conceptual language in order to reduce transformation error of IN as a concept in cognition to the level of interactions and ostensive activity. If one chose to represent objects in semantic classes, for example putting files into automatically generated directories pertaining to a semantic, this means that in order to deduce which class a new object belongs to there requires to be deduction of relationships between an information item and other information items. Such deduction amounts to relating a query, the new information item, to a set of documents which are the pre-existing information items, semantic classes and/or ontologies. Therefore it is proposed that search, at least on a conceptual level, is used extensively in the majority of semantics based operating system processes that begin with user interaction[23]. Indeed such operating system functionality is present in the recent

---

[22] Where opening the browser window corresponds to activating the OS.

[23] Whether it is present in underlying OS processes that do not necessarily initiate by user interaction depends on the design of the filing system structure, see [GBGL07].

versions of major operating systems (as elaborated in [GBGL07] for the Windows OS) and is seen to be an aim of future versions.

In the WWOS, search is not some abstract or separate concept but an integral part of every process where a 'process' in WWOS corresponds on a basic level to the process of visiting hyperlinks (moving between 'programs') and on higher levels it denotes application specific activities. Thus the traditional definition of a search as an individual program that is explicitly activated and interacted with is becoming increasingly inaccurate for describing the way search is used in normal operating systems. Moreover, as the form of an operating system is itself changing to become a system accommodating semantic/cognitive-like associations between items where search embedded in most operations, these issues require the traditional conceptualisation of search to perhaps be understood in a different way in order to approach modelling of retrieval algorithms from a general perspective including arbitrary types of interaction and result expression. Approaching IR from a more general perspective also pertains to the conceptual problem in IR that is addressed in the next chapter upon being related to the other research problems in the next section.

2.6 Discussion

2.6.1 Traditional and Ostensive Retrieval

The above discussion introduced some retrieval problems and approaches to tackle QFP. While the QFP is directly addressed, formally by retrieval techniques, the SUP is addressed through user training and usually any resolution of SUP is an indirect effect of usage experience. One of the sub-problems of QFP was that information need description as a query is achieved through a lossy process (Fig. 2.2) even when the user has some idea of it. RF techniques address this loss indirectly by assuming that the effect of the loss decreases over a search process through increased user activity and attribute this assumption to their underlying model and evaluation strategy. In doing this they inadvertently assume an underlying model for cognition which the OM generalises and makes explicit. This underlying model for loss of accuracy in expressing IN that is assumed in the OM is implicitly

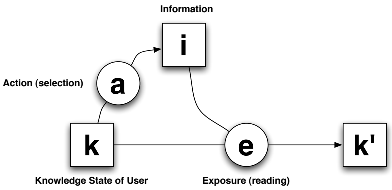


Figure 2.11. Knowledge State Flux in the Ostensive Model

inscribed indicated by the way probabilities of relevance are assigned to documents[24], thus ostensive retrieval currently does not directly address QFP/SUP.

In the cognitive model of OM (Fig. 2.11) at any one time the user has knowledge state **k** which is altered upon an action **a** from the user resulting in exposing (to a new path on the interface) their cognition to information **i** (documents), hence changing their knowledge state to **k'**. A further set of assumptions are provided in [Cam00] to relate IN/IN change with **k** / **k'** and user interaction with paths[25]. This is at most a generalised version of the cognitive assumptions that also underlie other RF techniques and specifically interactive techniques such as [WRJ02]. Such models generalise the 'reasoning system' in the agents involved in a HHIR process in which associations would be made between what had been seen/heard using methods from logical and other approaches. Instead replacing it with assumptions of how an interaction is related to an IN and how it evolves (ageing in OM).

Thorough study successfully applying formal cognitive models for purposes of interactive/ostensive search is lacking. Relative to the cognitive agents in HHIR current models of the user and also the system, which is supposed to imitate a second human agent, are insufficient. Apart from this it can be seen that the *interaction language* in the case of OM consists of clicking items on the path based interface, a very simple language. The question arises "What is the optimal interaction language" (relative to HHIR), what is the best way to communicate with users? In order to understand this one needs to analyse the nature of human cognition, the nature of a human cognition when in 'user' mode and the nature of a user, both in the traditional search context and otherwise[26]. In deducing

---

[24] For example, if one assumes much loss of IN accuracy then single user interactions mean less and therefore the probabilities of relevance of documents/terms are less affected except by multiple interactions with certain terms/documents. Hence assumptions of accuracy of IN can be represented in the system by the way probabilities of relevance are updated.

[25] State change is the most general way to describe the happenings of a search scenario, this is further elaborated in the next chapter; also IN change is assumed.

[26] As the next chapter elaborates this 'otherwise' does not exist once one re-defines search in an alter-

the optimal interaction language one must consider the balance between expressiveness, simplicity and retrieval effectiveness. Current CIR models do not allow such questions to be approached theoretically. Each retrieval application consists of an underlying model, an interface, and an interaction language but there is no overall framework/model to formally relate these components and reason between them[27]. This thesis questions the assumed cognitive models (user model), the retrieval models [Rob03], the way they are related (matching) and if it is viable to classify things this way relative to HHIR. More generally it investigates if there is a way one can 'ask these questions' formally, and how to then answer them since the current retrieval research frameworks do not easily allow it, hence our study into foundations. It seems that current models are indirectly approaching the problems associated with QFP and SUP which are deeply cognitive in nature. A fresh look from the first principles perspective is required to address these issues directly. Instead of detailing each insufficiency in terms of existing features of these models a top down approach is presented looking at cognition and what tools one needs to model it with respect to retrieval. The inadequacies of current models can then be *evaluated* on new grounds. The justification for this is made apparent by the proceeding discussion.

2.6.2 The Evaluation Problem

Evaluation described on an empirical level is the process of finding the values of variables representing properties of the system being evaluated. For example, in IR the system here is doing retrieval, the variables are that of precision and recall which are to be deduced for some instances of search. In natural language form it is proposed that without loss of generalisation, any process of evaluation is reducible to the process of answering the following question: *"How good is this system"*. The validity of this proposition is addressed in the proceeding chapter. If indeed it is a valid approach to represent the process of evaluation in this way then it is true to say that the purpose of performing evaluation on a retrieval system is equivalent to the purpose of answering such a question. At this stage an assumption is made, that the purpose of answering this question is that of determining

---

native way to include all communication.

[27] For example, it is not obvious without experimentation (and even then not formally explicit) if a particular type of interface complements (in terms of retrieval effectiveness) a particular document to query matching model or corpus.

the *quality* of the system being evaluated. It is assumed that quality[28] is an essential requirement for decision making processes whether it be human or artificial, thus it is critically important to deduce quality for the sake of continuity (of research). In the context of information retrieval systems, the ability to judge quality, perhaps initially only at a superficial level, allows systems to be compared. Specifically, a measure of quality would determine to what set a particular system belongs. If a measure partitions the set of systems into just two sets, then as research goes on and more systems are created, there requires to be further refinement of the notion of quality so that items within sets can be compared to one another. As quality measures improve and increase the discerning power of evaluation, richer techniques of representing the set of systems arise to meet the needs of increasingly rigorous methods of analysis[29]. Thus the process of refining the notion of quality is central to the evolution of the evaluation process and hence the overall research in the area. The prior discussion is a description of the notion of 'research' itself.

Repeatability An important characteristic of scientific experiments is that in many cases they can be duplicated exactly in another time and location. This is possible if each of the experimental parameters/conditions can also be reproduced[30]. In order to replicate an experiment, the defined parameters need to be well understood. In the natural sciences, experimental parameters often permit detailed, formal specification. If the parameters are well defined then experiments can be conducted with accurate control of each parameter/condition. These characteristics of parameters in scientific experiments allow computer simulation of experiments to become feasible[31], a potential cost benefit. One of the main difficulties in simulating an experiment becomes apparent when an experimental factor is not well understood. The experimental setup of a user study in retrieval evaluation therefore contains an inherent complication, the user. If the user was a controllable machine then one could note down user behaviours from one search experiment and duplicate them in another by initialising this machine with certain parameters. The user would then be a controllable experimental factor, since an instance of it could be formally defined in terms

---

[28] Without delving into philosophical discussion regarding the general notion of quality.

[29] Relevance and relevance judgements are one of many in this thesis.

[30] Reproducibility or replicability is an essential property of the scientific method.

[31] As an improvement in understanding the properties of the experiment is assumed to lead to an improved formal experimental model which in turn would lead correspondingly to an improved computer simulation of the experiment.

of behaviours. However, user behaviours are not yet expressed in such a formal manner and instead there are brief and informal natural language descriptions, i.e. the users are university students with moderate experience in searching. Ideally one would want to specify exactly, the character of each user, so the context of the experimental results can be better understood. There are several problems to evaluation done in the traditional way, firstly, conclusions to such an experiment are very subjective as they are limited to the scope of the test collection and to the context (factors influencing human perception of information) dependent viewpoints of prior human judgments. Secondly, there are no definitive ways, in general, to deduce why one system performs better than another since prior user judgments assess the system view of relevance; and are informal opinions of the whole system not specific formal reasons attributed to particular components. The latter problem is inescapable when using human judgments. In order to further understand experimental results from test-collection based evaluation (such as TREC [VB07]), one runs complementary live-user based experiments where users are given tasks to complete on a search system with effectiveness being judged using statistics on questionnaires and usage logs. An inherent weakness is that an experiment cannot be duplicated even if the same users are retained since their context changes. Thus the experimental results are not definitive. Indeed such problems are inherently due to the human factor, and will be referred to as the **evaluation problem**.

One aim of this thesis is to address the evaluation problem from foundations. It is claimed in Section 3.7that the question *"How good is this system"* is sufficiently representative of the evaluation problem for fulfilling this aim as to answer it thoroughly would entail analysis required to address evaluation issues in detail. Justifications for this claim may be realised by considering firstly that in order to answer the proposed question, its constituent parts must first be defined; thus the meaning of 'system', 'good' and the phrase 'how good' must be expressed. Further, the issue about the present tense of the sentence due to the 'is', must be addressed to specify if the evaluation results apply for one time or are applicable over multiple periods of time. The assumption is that the constituent parts of this question are sufficiently variable to generally represent (at least) the majority of the evaluation problems for retrieval. In summary the importance of this question is that it is central to the decision process used to attach interpretations to the subject of the question, which in this case are retrieval systems. Without the ability to decide between systems there can be no progress in applying research. The question is claimed to be appropriate in representing evaluation on a very general level, and was chosen as a starting point for addressing the foundations of evaluation in retrieval as elaborated in Chapter 3.

2.6.3 Unified Retrieval Model

Consider the fact that in order to judge a system user opinions would need to be considered, even for simulated users. A user carries context, which include their environmental factors for a real user and particular definition for a simulated user. A simulated user is strongly coupled to a set of user-system interactions, and an interface. The interface and the user-system interactions components are usually only specified in informal natural language expressions in search. Overall, there is no way to formally specify an IR experiment in its entirety. As a result there is no way to formally reason about evaluation results with respect to the interface and interactions.

As user studies are more empirical in nature than theoretical there is uncertainty and complexity in the definition of research problems and methodology; and in the interpretation of experimental results. It is proposed that the reason for this is the lack of a formal theory unifying the methods of expressing and reasoning about the different search components. Absence of such a theory limits our ability to compare in a sound way the research problems, methods and results in information retrieval. A successful unified theory would be required to amalgamate the different representations of a search component, such as the vector space and probabilistic models at the matching/decision level. It would also require integrating the manners in which a component is described and studied. Ideally this theory would allow search elements to be formally described and reasoned about in relation to one another. The search elements described as the interaction language, interface, decision/matching mechanism and data corpus require to be represented by a common formal theory.

Traditionally, only the document, query and matching models (Fig. 2.1) are formally specified and admit several formal specifications. For example, the association between a document and query termed relevance can be represented in terms of logical implications, conditional probabilities or inner products in a vector space. Currently there are no unified frameworks for theoretically comparing between different representations in terms of effectiveness. These problems with rigorously characterising users, interface, user-system interaction, and the inability to compare different formal representations where they exist, hinders research as it severely limits theoretical conceptualizations of search scenarios and deductions therein. In comparison the simple physical system allows many degrees of freedom for devising hypothetical extensions to already specified scenarios, such as extending Fig. 2.3 with two balls and many walls. Definition problems apply not only to specifying a search scenario but also to the relationships between the IR research field and neighbouring fields as stated in the introduction.

Overall, the specifications of experiments in IR lack formal expression. This restricts the capability to formally reason about evaluation results with respect to these search elements. Traditionally, only the document and matching models have formal specifications in retrieval experiments. In comparison, the method of specification in the sciences, practical physics for example, is formal in more aspects. Thus, the way a retrieval experiment is represented creates ambiguity due to the informal natural language expression of several of its parameters. However, in general IR research, these retrieval elements are represented, written about and thought about in this way, out with experimentation. This state of affairs in IR is reminiscent of a similar situation from the beginning of the 20th century when Hilbert and von Neumann proposed to axiomatize branches of mathematics and physics [Cor97], suggesting a systematic approach to expressing theoretical and experimental claims and reasoning about them. The consequence of this drive towards systematic specification was a deeper understanding of the structure of the problems being specified and limitations of the formal specification system itself; resulting in theorems of computability and incompleteness. The formal specification system that later emerged allowed computers to simulate mathematical models, and anything that could be in turn modelled by the mathematics, such as physical and chemical processes. In order to make the experimentation in IR more like that in the sciences and make good use of simulations, a formal system for expression of search scenarios is necessary.

Formally defining a retrieval process, at first glance, seems implausible due to the complexity of the user element. The mathematical modelling of physics did not begin by attempting to describe the whole universe but by instead describing much smaller phenomena. Hence given a generalised specification language and a method of axiomatization one could begin by formally representing very specific experiments with very specific user behaviours and gradually building a database of a multitude of possible retrieval scenarios. Although a real user may never be exactly specified, the formalism could reveal facts about the specific limitations of user simulation just as the theory of computability revealed restrictions inherent in modern day computers. There is also the potential to understand on a deep level when one can do simulation and when they must resort to user studies.

2.6.4 Method and Representation Spaces

In order to understand the breadth of a formalisation task for IR the concepts representation space and method space are introduced. The representation space for a search component is defined as the collection of all specifications of members of the element. For example, vector spaces are one way to represent the matching and document models; therefore they are part of the representation space for these retrieval elements. The interface can be represented by functional descriptions and diagrams, which form part of its representation space. The method space for each element defines all logical combinations of the specifications in the representation space. These logical combinations are precisely the way in which the vocabulary, that is, the representation space, is used in expressing the element. For example, the part of the method space corresponding to a vector space representation (a part of the representation space) consists of a mixed (natural language with mathematics) description of how to perform a cosine similarity operation and express its interpretation. Alternatively, the representation space corresponds to syntax for expressing a retrieval element whereas the method space corresponds to the semantics of the representations, Fig. 2.12 illustrates this. Restating the message of the previous sections in a new way, the problem with IR in terms of this type of characterisation is the multitude of such spaces and there being no way to compare/contrast between them. Therefore, in terms of these concepts it is suggested that the aim for IR be to find for each search element a new representation space that can accommodate several of the significant representations

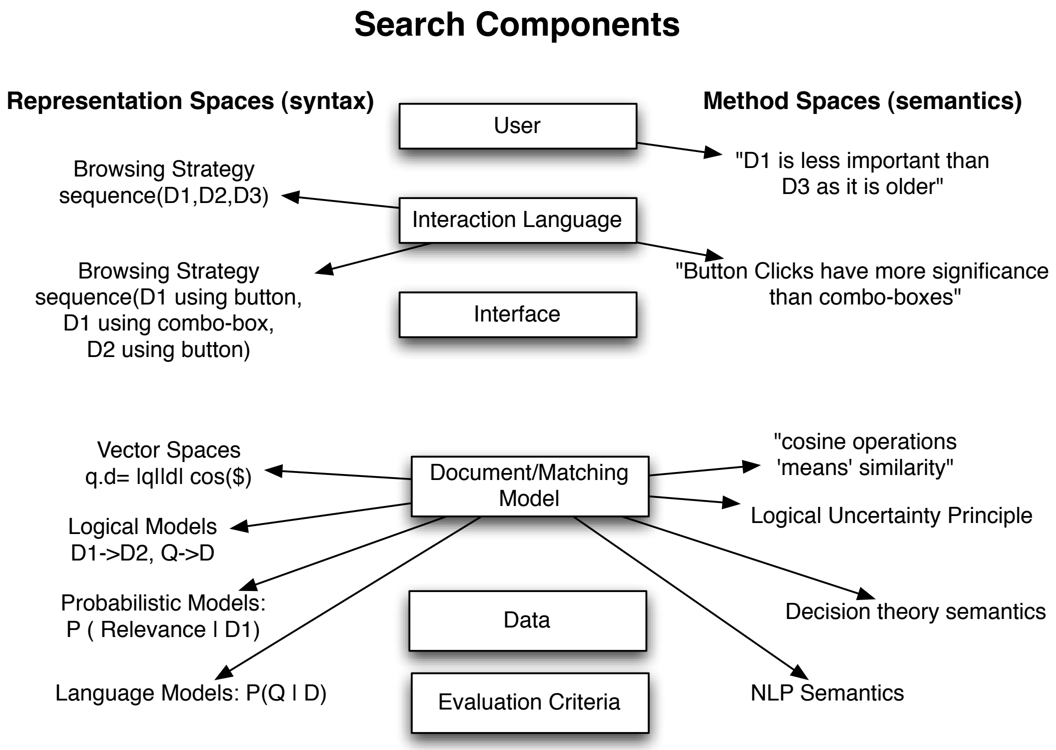


Figure 2.12. Method and Representation spaces of search components

that are around[32]. A general theory for each component is then required to form a complementary method space to allow reasoning about the different representations in relation to one another[33]. Such a theory would have to agree with the semantics defined by the prior method spaces of individual representations so that the new theoretical constructs exhibit consistent interpretations. The idea is to map each of the retrieval elements to a more general space keeping the mapping as isomorphic as possible. Following this the task is to find a representation space and method space which can accommodate all of the retrieval elements. This final representation space is denoted the unified language and the corresponding method space the unified theory[34].

---

[32] For example, the Hilbert space as suggested in [vR04] that accommodates the vector space, probabilistic, and logical models that are used to model the matching component (among other components) in an IR system.

[33] If the representation space for the matching components are the Hilbert space the general theory corresponds partially to methods of use of the Hilbert space from QT. For example, a representation of a probabilistic retrieval model can correspond to *density operator methods* (see Chapter 4) for working with probabilities whereas a vector space model would be associated with geometrical methods typically used in analysis of QT systems. As both methods are part of the same QT there are natural ways to compare them. However, the results of such comparisons may not immediately have a retrieval specific meaning from which useful hypothesis about models can be generated.

[34] This thesis provides initial work for a unified language and theory, see Chapter 5.

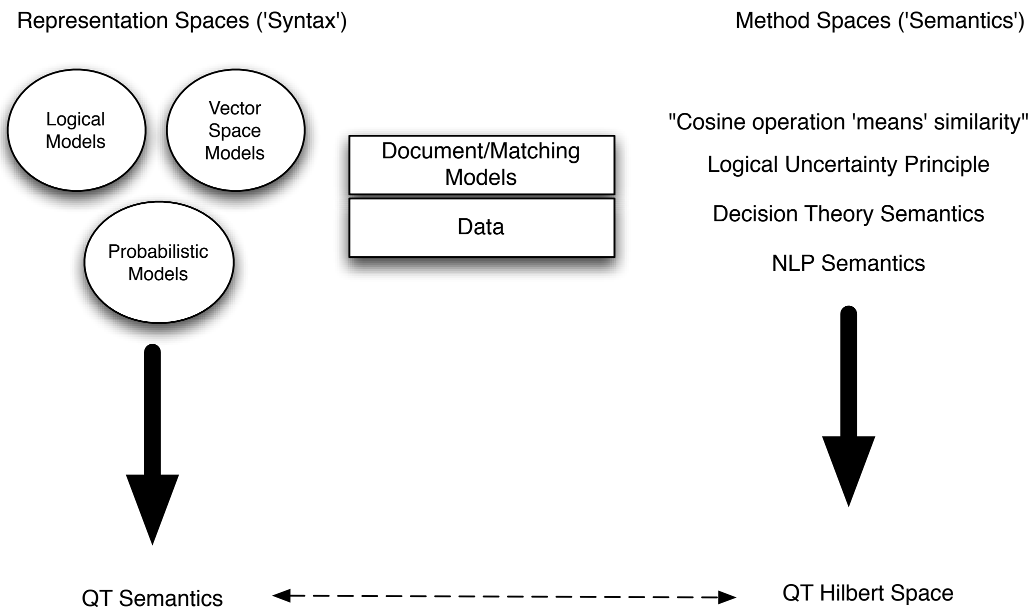


Figure 2.13. Mapping of Method and Representation Spaces

On analysis of the matching and document modelling elements it was found in [vR04] that Quantum Theory (QT) is able to accommodate the main representation and method spaces, such as probability spaces, vector spaces and logical models which are traditional for specifying the decision and matching elements Fig. 2.13. As elaborated in this chapter it appears that the theory can also account for many cognitive phenomena and hence accommodate the representation/method spaces describing the semantics of very specific user behaviours. If the user behaviour (a subset of user cognition) is mapped to the QT framework and inherits the QT method space. Previous work in [Xie02] defines some of the common representation and method spaces of the interaction language, interface and some user behaviours. In the new method and representation spaces the user component would be expressed using the vocabulary of the inherited specification language as a quantum-physical process in terms of quantum state changes, measurements and phase-shifts. Decisions made by a retrieval system, its data representation, and matching models would be expressed using the same set of theoretical constructs and rules when mapped onto the QT framework. The evaluation element once mapped to the QT framework can be interpreted in a manner similar to the way a measuring device is understood in the same framework. The mapping of elements from their initial method and representation spaces to the generalised representation offered by the QT framework is difficult as in general, the initial spaces are not well defined. Ultimately analysis of the large number of method/representation spaces is required especially for those elements not exhibiting strict formal definitions, such as the natural language elements in Fig. 2.12. The aim in such analysis would be to find commonalities in the representation and method spaces of these elements which would aid in deducing the theoretical constructs for representing the elements in a QT framework, so that the commonalities remain in the new representation.

One of the main aims in mapping retrieval to physics would be to attain the ability of doing formal analysis of search and to acquire the experimental strengths of a QT framework. The ability to formally specify retrieval elements can be used to define retrieval scenarios and their change over the period of a search session. Hypothetical scenarios or simulations can then be formally specified and used to reason about the corresponding simulation results. Simulation in IR has helped with the analysis of specific retrieval scenarios [WJvRR04] but there are no guidelines indicating how the current simulation techniques could be adapted to general scenarios. It is proposed that formal specification would provide a deeper understanding of the relation between the simulation parameters and the results. The formal reasoning system exposed could be used to decide how simulations are related to one another and the validity of inferring (without running a simulation) the results of one simulation from that of another with the intention of reducing costs. An additional foreseeable advantage of mapping a retrieval simulation onto a quantum physical process is that there are instances where problem (expressed in some QT framework) can be solved more efficiently on a quantum computer than on modern day (classical) computers [Gro96]. This may benefit retrieval evaluation if quantum computation were to become feasible in the future.

2.6.5 Abstraction of Research Problems

The research problems discussed above can be grouped into three broad categories. Firstly there is an inability to verify experimental results in IR and other issues related to evaluation, these shall be collectively known as the **evaluation problem**. Secondly, the evaluation problem is related to inability in formally specifying users, which will be denoted as the **user problem**. Finally, the **definition problem**, referring to the problem of not having one

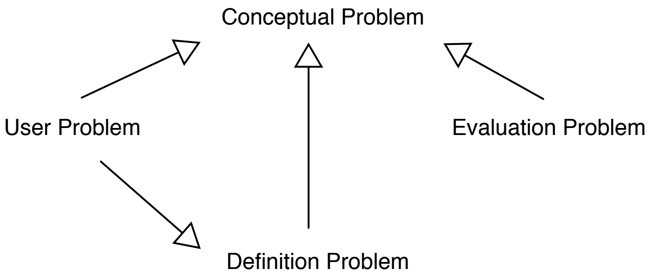


Figure 2.14. Dependence of Research Problems

representation/method space from where to relate all other spaces, denotes the issues with formally defining all components, their inter-relationships, theoretically reasoning about them and the relationships of IR as a field with other fields. Resolution of the user and evaluation problems depend on the resolution of the definition problem. A fourth problem yet to be elaborated is the **conceptual problem** which is addressed in the proceeding section completing the relationship between our research problems as illustrated in Fig. 2.14 where the arrows denote dependence. The dependency between research problems is not formally provable on a general level, the figure serves only to illustrate reasonable relationships as per common experience in IR research. It is at the same time interesting and unfortunate to note that due to the definition problem it is difficult to present without ambiguity these above research problems in a formal way.

The four categories of problems appear to sufficiently abstract the issues faced in our initial research, which in hindsight attempted to resolve the definition problem. Initial research in [vR04] can be said to partially address the definition problem by suggesting an unified theoretical basis for comparison between different types of document, query and matching models expressed in various formal specifications. The work in [vR04] showed that mapping of three types of these models, the vector-space, probabilistic and logical models to the mathematical formalism of Hilbert spaces in the way employed by quantum theory results in a single framework in which one is able to theoretically compare among models, providing greater opportunities for formal analysis than previously available.

One important aspect not elaborated in [vR04] was that of modelling relevance feedback in the QT formalism. In an attempt to model relevance feedback in the Hilbert Space formalism of quantum mechanics a new set of problems were faced and interesting questions raised which collectively suggest novel inquiries about the nature of IR; these are elaborated in Chapter 4.

2.7 Summary

This chapter briefly discussed traditional retrieval approaches, the concept of the user, the system, their interaction through relevance feedback and specifically through implicit feedback in an ostensive retrieval system. It was concluded that the problems in IR that prevent it from full scientific status pertain to evaluation which in turn pertain to lack of an unified formal theory for defining the multitudes of search concepts. The definition problem was then proposed to depend on the conceptual problem which asks 'what is search?', as if the traditional search system now admits alternate definitions and the method of interacting with search systems become less distinguishable from general computer use then what are the boundaries between a search task and any other task? The next chapter aims to generalize the concept of search in order to address the conceptual problem. Then it is shown in Chapter 4 that mapping form retrieval to QT is difficult and only useful in certain situations. Aspects from QT are then taken to create a new representation and method space in Chapter 5, a middle language between QT and IR combining characteristics of both for addressing the definition problem.

—He who sees Me and **knows** *that he sees Me, does not see Me*

Muhyuddin Ibn Arabi (1165-1240), Futuhat IV p.55

3

On Reduction to the Higher Ostensive

3.1 Introduction

This chapter discusses the proposal that information retrieval regardless of its various forms can be reduced to a particular representation which is superficially similar to ostensive retrieval [Cam00] in both its visual representation and in supporting the ostensive philosophy. Toward the end of the chapter some fundamental questions about IR are addressed in dialectic form elaborating on conclusions drawn from the following initial section with regards to user simulation. Initially the idea of state and state change is introduced for understanding a search process.

3.2 Notion of State

Firstly notice that given a search interface that accepts a natural language query one can deduce that there are only a finite number of possible queries, and therefore a finite number of possible results given a search model (including models pertaining to data, matching and interaction/interface). In practice, relative to the large number of possible queries there will only be a smaller number of frequent queries and results that correspond to them. Secondly, even with a complex user interface, there are only a finite number[1] of possible interactions and similarly in practice there is a much smaller quantity. Whether a user types a natural language statement or interacts in some other way to indicate their query, there are overall a finite number of possible moves. Each move corresponds to a not necessarily unique result (response). At each point in time during a search process its is said that *the search is in a particular state (of affairs)*. A user then chooses a *move* whose act causes the system to change its state into a new state. The creation of this new state corresponds to the effect of the search-interaction. The state of an object is an abstract device representing the object from which a set of truth statements about the object, not necessarily all true statements, can be derived[2]. An abstract device is synonymous with 'a theory', or conceptualization, and is necessarily always incomplete in that, for a search

---

[1] Hypothetically it may be possible to devise an interface with a randomly changing look and therefore due to the large number of possible interactions it could be useful for analysis to consider the interface as one allowing infinite interactions.

[2] State defined this way is compatible with the quantum theoretic notion of the state of a physical system where a system is completely described by its *state vector* (see Section 4.2 and [Gri02, NC00]).

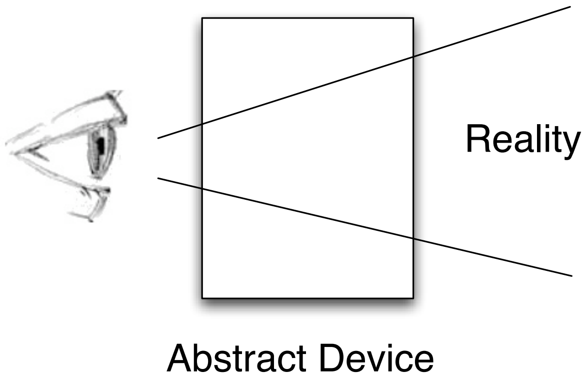


Figure 3.1. Theory as an Abstract Device

process there will be a statement about the process whose truth value cannot be deduced by the abstract device (Fig. 3.1).

A search state contains a set of truth statements about a search process, there is further structure to these statements. In traditional IR, the statements are (implicitly) grouped into either a user state or a system state when they exclusively refer to either the user or system components respectively; thus they are sub-states of the search state. Statements about a user's search interests would be associated with the user state while assumed truths about the relevance of documents would be associated with a system state. Since the user and system continually influence each other during a search, there would be significantly fewer statements which are exclusively associated with an agent than statements involving both agents. The groups into which truth statements are placed form an *ontology*[3] for the search process, the traditional search ontology consists of the dual: user and system[4]. The purpose of the traditional ontology is to answer the question "what are the *knowable* things about a search?" in terms that are useful. Research in IR has taken the user-system dualism as a normative ontology as it is useful from the practical viewpoint of creating 'the system', that is, search software that requires to know something about a user involved in a search process with itself. This practical viewpoint necessitates that the system be a composite entity as its interface with a user is separate from its internal workings. Similarly a user's cognitive activities are not necessarily expressible through the set of possible interactions provided by the system interface, therefore the user state consists of at least two contrasting

---

[3] For the purposes of this thesis ontology can be taken to mean that *about which* things can be known.

[4] This means that traditionally all that can be known about search either pertain to the user or the system, i.e. is either a user or system property.

sub-states, a cognitive state and an expression state. The expression state is associated with statements about observed user behaviour while the cognitive state is associated with statements regarding reasons for this behaviour. In this way a search state, which refers to a set of truth statements about a search process, is structured according to a useful ontology and consists of sub-collections of these statements which is said to correspond to *sub-states*[5]. What is knowable about a system, an user or the search process in general[6], depends on the agent that is inquiring. From the point of view of a system, it derives some statements from observing user interaction which are traditionally known as statements about 'relevance'. The system typically has minimal and highly uncertain knowledge about the cognitive state of the user. The user's knowledge of the system varies according to their understanding of the inner workings and information domain of the system. A researcher observing a user searching using his/her software can potentially know things about the search process that the user or system cannot. Relationships between statements knowable by the user, system, and researcher are explored next.

3.3 Agents, Knowables and Incompleteness

It is argued that there are at least three agents present during search, the user, system and researcher. Each agent has limited knowledge about the other. There are also a set of knowables which are not necessarily all known. In Fig. 3.2 an agent is shown as a set of statements, where the set $s_{1.x}$ denote statements associated with the interface the system exposes to the outside world whereas $s_{2.x}$ are the agent's internal details. Similarly $\widehat{s}_{1.x}$ denote the user's exposed properties, their interactions with the outside world and $\widehat{s}_{2.x}$ are statements pertaining to details about their cognitive state. There are also truth statements associated to the search process instead of any one agent these are denoted $\widetilde{s}_{1.x}$. The source of an arrow denotes the *knower* and the destination denotes the statement that is *known*. In the case of Fig. 3.2 both the system and the user know one statement about the exposed part of the other. The user knows the general statement $\widetilde{s}_{1.2}$. The researcher knows the internal workings of the system, one general statement and all statements pertaining to

---

[5] A substate is therefore a factorisation of a state; unless one knows the power set of all statements in which case a sub-state is a subset of the set of all subsets of statements.

[6] There can be statements associated with a search statement independent of its participants.

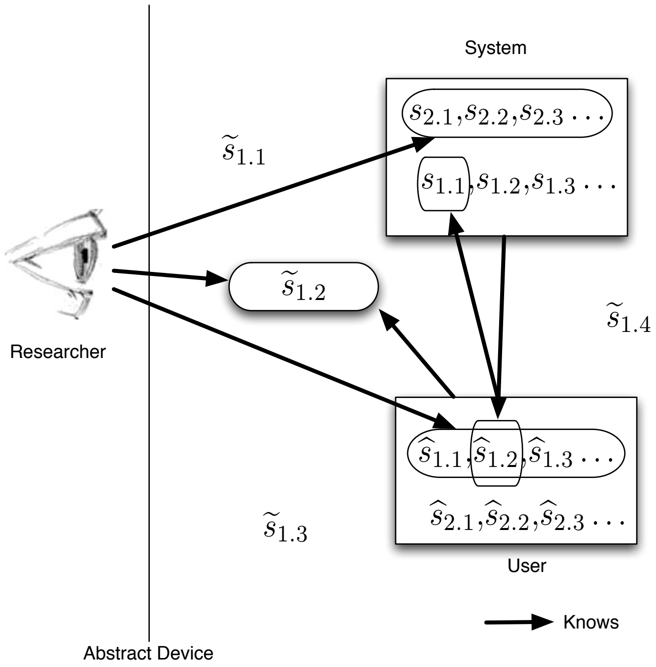


Figure 3.2. Knowledge of Agents

user interactions ($\widehat{s}_{1.x}$). A traditional search process in terms of Fig. 3.2 is an activity in which the knowledge of all three agents are altered since the user learns more information, the system learns about user interests and the researcher learns both about user interests and system performance. A traditional search goal from the user's perspective would be to learn something relevant from the system and for the researcher to learn the truth of a general statement ($\widetilde{s}_{1.x}$) such as that pertaining to precision/recall values of queries formulated by the user. As the traditional system is deterministic its goal is fulfilled by it simply running in the preset way. If one adds temporal conditions onto the system such as requiring it to seek particular interactive behaviour from a user, for example, requiring the user to rate query results favourably on the fly, then this becomes part of the system's search goal. All such goals for agents in search pertain to the objective of altering the knowledge of the agent.

Statement $\widetilde{s}_{1.2}$ is known by the user and researcher, an example would be a statement among statements pertaining to the progress of a pre-formulated search task such as that of browsing some specific documents as done in traditional user-based evaluation. There are general statements such as $\widetilde{s}_{1.1}$, $\widetilde{s}_{1.3}$, $\widetilde{s}_{1.4}$ which neither of the three agents know or ever know throughout a search. For example, $\widetilde{s}_{1.4}$ could be a statement about the current search process which indicates that given all its properties the process is the most effective of all such processes ever executed[7]. Whether such a statement can exist, is in itself a ‘research problem’ that may have never previously been investigated thus the researcher (introduced in Section 3.2) has no way of knowing about the truth of the statement during the search scenario. Moreover, given that the researcher does not *know about* $\widetilde{s}_{1.4}$, it is not necessarily the case that the truth value of this statement is either deducible at all or deducible within a finite time frame and with finite resources, also perhaps the statement is only partially deducible and perhaps only with probabilistic certainty.

In the point of view of a writer writing about IR research there are three agents but from the point of view of a reader of that research there is at least four, the writer doing IR research, the researcher being written about who is present at a search scenario, and the user and system in the search scenario. A research reader can view the research writer in the way the research writer views the search scenario research agent. The research reader investigates the correctness of the research writers compositions. Similarly a report about the research writer from a researcher reading the research can itself be investigated by another researcher and this can continue *ad-infinitum*. If one defines search more broadly so that it not only denotes a user-query-system based activity but also any activity by which an agent’s knowledge is altered then the research process is inherently a search process a notion that is etymologically indicated by the word research/‘re-search’. Using this broad notion of search Fig. 3.3 shows the research process ad-infinitum labelling it ‘Re*search’ where ‘*’ denotes indefinite application of search[8] indicating that the ‘Re’, the research of search, occurs zero, one or more times. Given that the current thesis is research it is necessarily restricted according to the concepts it explains. Objective research about search is therefore limited due to its self-referential nature. Further limits to search research are associated with the modelling and simulation of the user agent as discussed in the proceeding.

In traditional IR research, the ontology of search processes discussed in publications is from the point of view of the researcher. Can it be perceived it in another way, for example, what

---

[7] For this to be the case there would be some time factor or memory associated with $\widetilde{s}_{1.4}$ to indicate that it knows the set of all solutions.

[8] In this context the Kleene star operator [Rog87] denotes repetition of the search process indefinitely which in turn is taken to denote a research process; thus it is proposed that research can be seen to consist of a not necessarily finite number of search process.

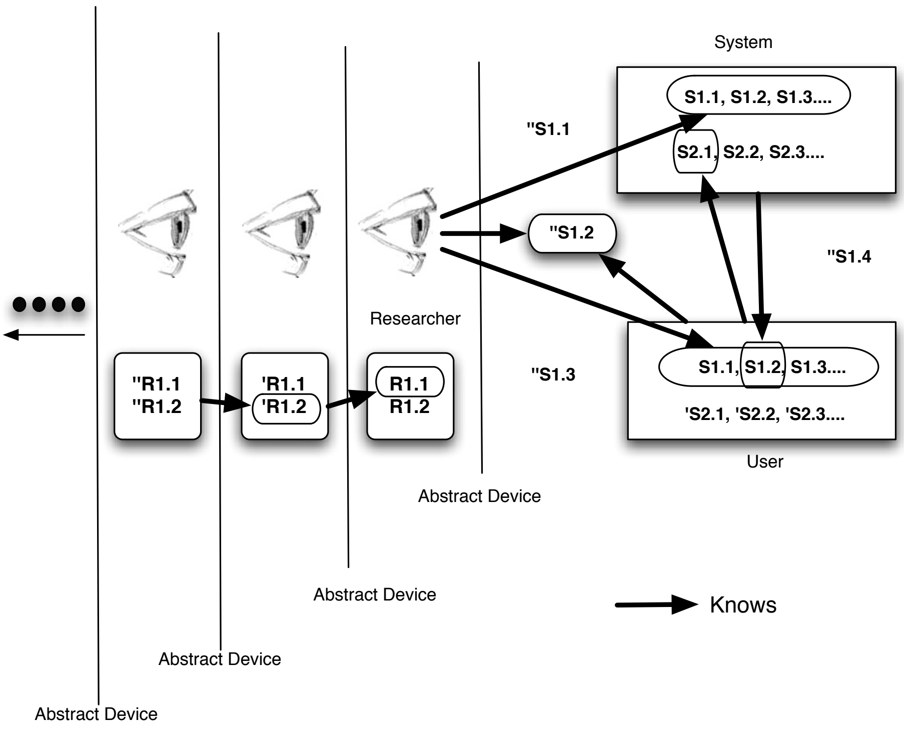


Figure 3.3. Re*search

about as a simulator, or as someone observing search as interactions between two systems? Can one think of the elements of a search process in terms of states or are there other forms as in traditional retrieval that are useful for expressing search concepts? These issues are addressed in the next few sections. Given the *knowable* things about a search process, the next step is to deduce what is known at some point and how it is known, these are epistemological issues of a search process. In traditional IR, the user interest is a sub-state of the user state[9] and corresponds to a set of truth statements some of which are derived through analysis of user interactions. The *epistemology*[10] of user interests is embedded in the model that interprets user interaction. The system's view of a user's interaction is traditionally represented by a set of keywords associated with the user interactions which are assigned numerical values to denote their importance. Keywords with high to low numerical values form a descriptive interpretation of user interests with decreasing accuracy. Prior

---

[9] The user state denotes all knowledge about the user; this is further elaborated in Chapter 5.

[10] Interpreted as (1) the dynamics (how it changes) of user interests and (2) the way user interests are known or ascertained. In general, for the purposes of this thesis, the epistemology of a search process pertains to the way statements are known about the search. Epistemology is defined in this way due to the premise of this chapter that all research and *ways of knowing anything* can be perceived as search. Thus, the production and acquisition of knowledge (by an agent) is said to either *done through search*, or *specifiable in terms of search*.

discussion addressed the 'statics of search' concerned with ontology; the next section discusses how the knowables are known over time, thus addressing the epistemology and dynamics of a search process.

3.4 The Ostensive Philosophy and State change

In the traditional search interface the user interacts with the search box and search button. A more generalized interface is that of the ostensive retrieval interface of [Cam00] which presumes that the user's information need is defined ostensively through pointing at information items. An ostensive definition is that which conveys the meaning of a term by pointing out examples, further elaboration is presented in the quote from Wittgenstein:

> So one might say: the ostensive definition explains the use–the meaning–of the word when the overall role of the word in language is clear. Thus if I know that someone means to explain a colour-word to me the ostensive definition "That is called 'sepia' " will help me to understand the word.... One has already to know (or be able to do) something in order to be capable of asking a thing's name. But what does one have to know? Ludwig Wittgenstein, *Philosophical Investigations, Sect. 30* [Wit01]

Can one then represent all interfaces in a unified way since the user is just pointing or 'ostensively defining' concepts in all search interfaces? This can be addressed by noting that a user interaction influences and thereby indirectly causes a change of state in other agents. The system's knowledge of the user's state is changed due to the new evidence about the user state available through user interactions. Similarly as the system reacts, traditionally by outputting results or more generally by simply capturing the interaction, the user's knowledge about the system's state changes. For the researcher who usually knows more about the system than the user and also usually knows the user better than the system, his/her knowledge of the user and system states is also changed. On the ostensive interface search proceeds by the user selecting documents upon which further related documents visually branch out from the previously selected document. A similar visual diagram is used to denote state change where possible future states branch out from

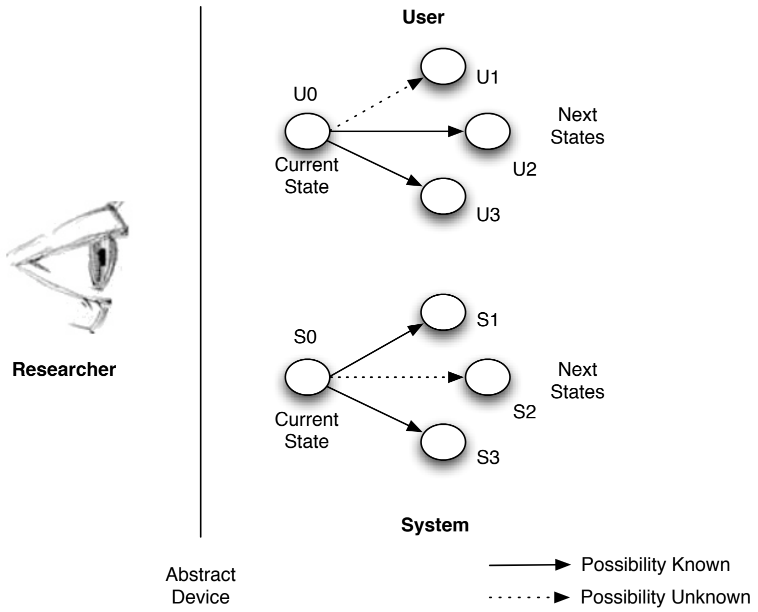


Figure 3.4. Researcher Perspective

the current state. The state change diagrams in Fig. 3.4 denote the researchers knowledge about the system and user's states. Similarly the user's and system's knowledge for other agents are depicted by Fig. 3.5 and Fig. 3.6 respectively. Each diagram is done from the point of view of a hypothetical observer that knows all states an agent can change to at any point such that it knows if a particular agent knows possible future states of another agent. The dotted arrows represent the case where an agent does not know about a possible future state of an agent at some point in time, for example in Fig. 3.4 the researcher hypothesises U2 and U3 as the only possible future states for the user. The researcher will only deduce the possibility of U1 through further observation of user interaction, that is, by acquiring more knowledge. In the case of a researcher aiding the user during a search process, the user would be interacting with the researcher and thus would be knowing or wanting to know information about the researcher. A more common real-life scenario is of a librarian or a blog aiding a user's search instead of the researcher. Similarly Fig. 3.5 also depicts the possibility of a researcher (or other agent) aiding the system during search; in the simplest case this could refer to live-tuning a system.

In general at a particular time during search the agents are missing information about (1) possible future states of other agents and (2) the actual future states in which other agents will end up. The researcher's knowledge about the system is a possible exception to this

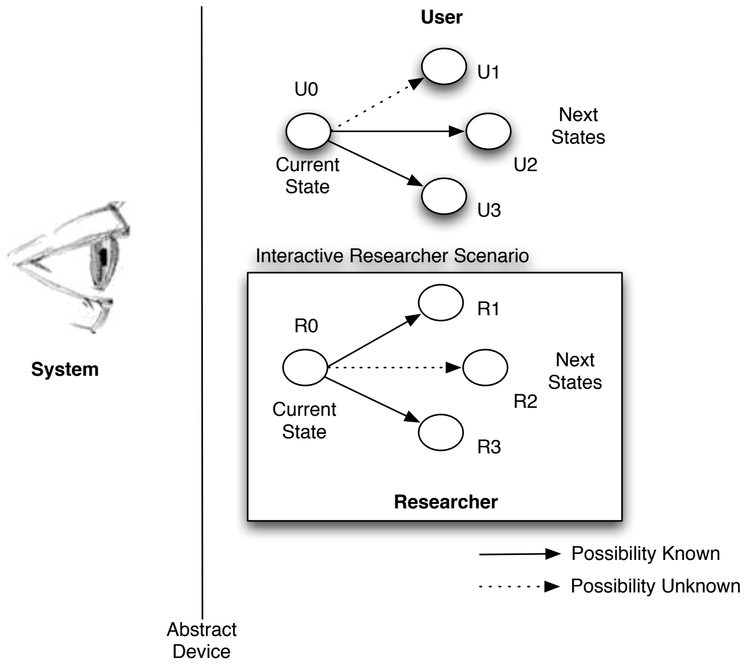


Figure 3.5. System Perspective

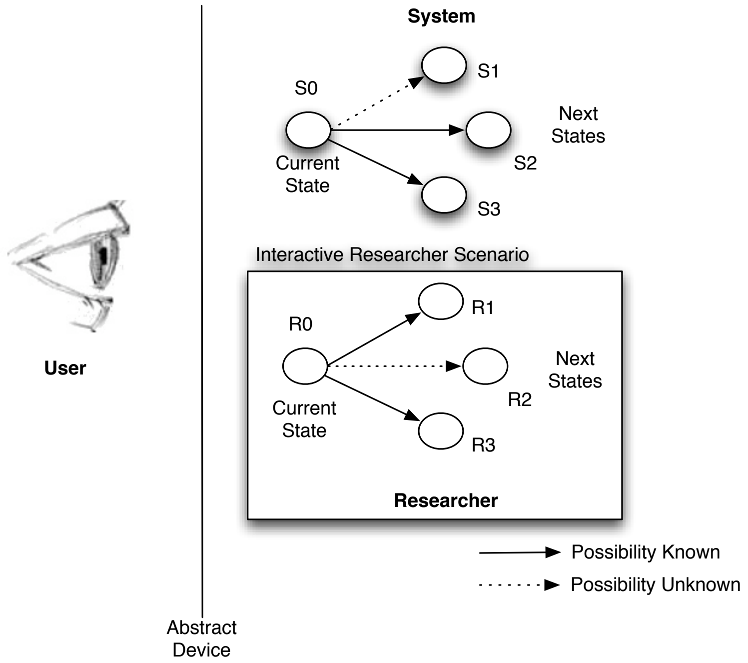


Figure 3.6. User Perspective

rule; in most cases any knowledge of changing states of an agent can only be approximately determined by any other agent as per traditional retrieval (i.e. user modelling). The missing information of agents or knowledge gap, a superset of the traditional semantic gap[11],

---

[11] The *semantic gap* is a particular of a *knowledge gap* since it refers to a lack of knowledge about the meaning of a data item (such as an image or video). The knowledge gap in turn is due to the *social gap*

is addressed through gaining knowledge through further observation of agent behaviour and/or internal deduction (known as *internal flux* in [Cam00]). Agent behaviour is observed when actions happen thus they occur ostensively. Ostensive definition is not only what the ostensive retrieval framework presumes about user interaction and requires of an user to do, but it is also, on a general level, what any agent necessarily requires of other agents. In general, due to lack of knowledge of agents full deterministic understanding of agent behaviour is restricted instead it is required that all states that are substates of the search state to be ostensively defined. Ostensive definition is therefore central to the epistemology of search.

3.4.1 The Higher Ostension

The Research Oracle

The research oracle is depicted in Fig. 3.7, its perspective is presumed in Fig. 3.4, Fig. 3.5 and Fig. 3.6 where the agents' knowledges are denoted relative to that of the oracle's[12]. The research oracle knows all possible states for other agents at any point in a search but it does not know for certain the state the agent will take next[13]. Thus it also presumes ostensive definition of states for other agents. As soon as the current state is established the research oracle knows all the details about it and possible future states but how is the current state chosen? Given a set of possible states the next state is chosen by an *external oracle* as shown in Fig. 3.8[14].

(introduced in Section 1.2), thus it is assumed that that the knowledge gap is a particular case of social gap with the idea being that if the creator of data and searcher of data knew one another then there would hypothetically be (upon HHIR) minimal (or no) knowledge gap between them. At a more generalised level it is said that that all these gaps are cases of an *experiential gap* which is taken to be the 'highest level gap' and is associated with the generalised concept of information need (see Section 3.7.4).

[12] To be able to refer to the complete knowledge of agents it implies that the referrer knows this knowledge, thus the oracle is said to contain all the knowledge the agents it observes contains; therefore the diagrams denoting the knowledge of agents imply that they are from the perspective of the oracle who completely knows these agents.

[13] It is introduced here as a necessary precursor to the concept of a higher ostension; the general utility of this concept is an open problem, however, some pointers are given in the proceeding discussion

[14] Recall that the research oracle knows everything about the search and agents except what the upcoming state will be.

System and Research Oracles as “User and System”

On a higher level, the research oracle is to the external oracle as the system is to the user since the research oracle must wait for the external oracle to act in a way similar to the way the system waits for the user to act. The user initiates a search and is central to the purpose of existence of the search process. Similarly the external oracle creates state change by selecting new states. Further, it can be presumed that in the most general sense if one is to assign a purpose to any process it would be to change state, therefore the external oracle is central to the purpose of the search process. The higher ostension refers to this set-up where way the external oracle causes state change in the search.

Simulation

Could the research oracle ever be a practical device instead of an abstract concept? Consider the case where instead of a user and system there are two systems then the research oracle, that knows all possible future states of a search state including states of agents, is a simulation program or device. If one of these systems can be said to be simulating a user then the process is a search simulation. The perspective of a researcher observing a simulation device as shown in Fig. 3.9 is a scenario where through interaction with the simulation software that the researcher is able to know all possible future states for the search. How does simulation change the way things are known by agents? For the system and simulated user the ostensiveness[15] of events is variable depending on what it known to it through its model (program) depicting the system and search process. For the researcher what is ostensive or unpredictable from the start is the final result, as if the final results are known then a simulation is not necessary. The researcher is able to know at least the boundaries of the simulation results as embedded in the design of the simulation. The knowledge gap is then *localised or reduced* (in that it only consists of the lack of knowledge of particulars of the modelling of a real user) to the background research one conducts to create the simulated user as depicted in Fig. 3.10. While background user research may

---

[15] Ostensiveness here means predictability or uncertainty in any predictions it makes. However, uncertainty or unpredictability is only an effect of an ostensive act as seen by an observer. In general, ostensivity of an agent is a measure of the knowledge the agent has of another, thus the ostensivity of A is $Ost_A(B) = f(\textit{What A knows about B})$.

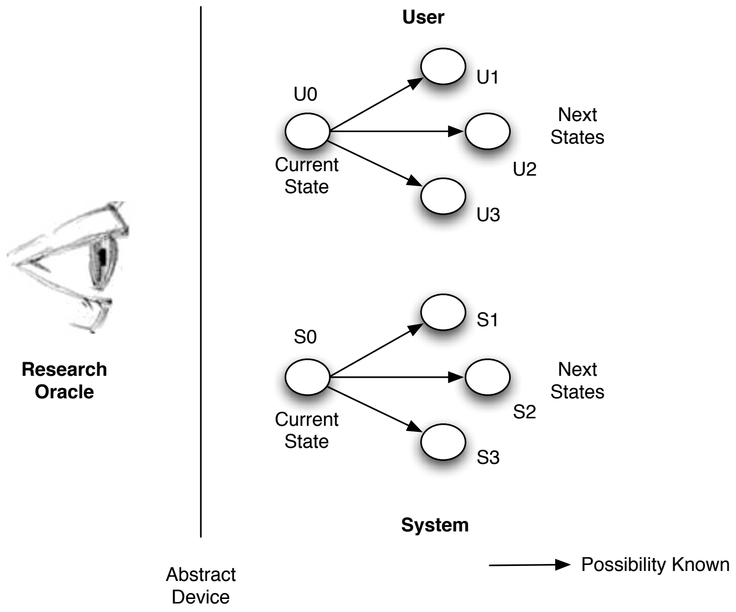


Figure 3.7. Research Oracle's Perspective

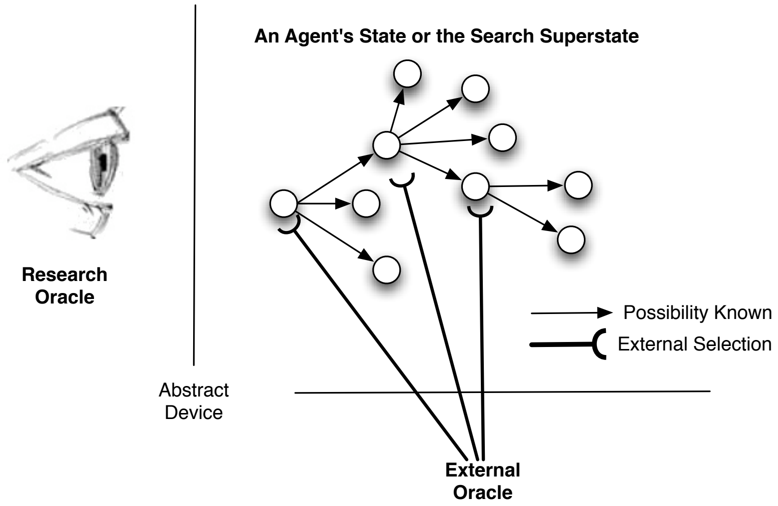


Figure 3.8. External Oracle's interaction with Search

involve complicated processes including search processes. The user models generated can be used repeatedly in simulations. In the traditional context this is equivalent to employing the same users between several search scenario based investigations during different time periods all behaving in the same way and with little or no cost associated to all but the first such investigation[16]. Therefore, simulation affects IR research methods and it can potentially have significant effect on the management and administration of research.

---

[16] This depends on a marketing strategy for user models, for example, in the case the models are to be shared.

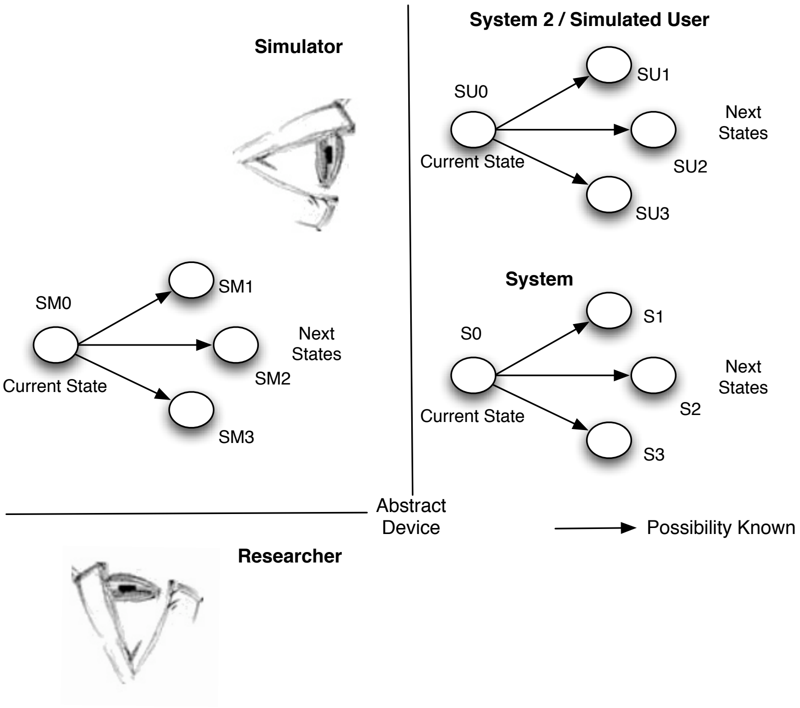


Figure 3.9. Researcher perspective in Search Simulation

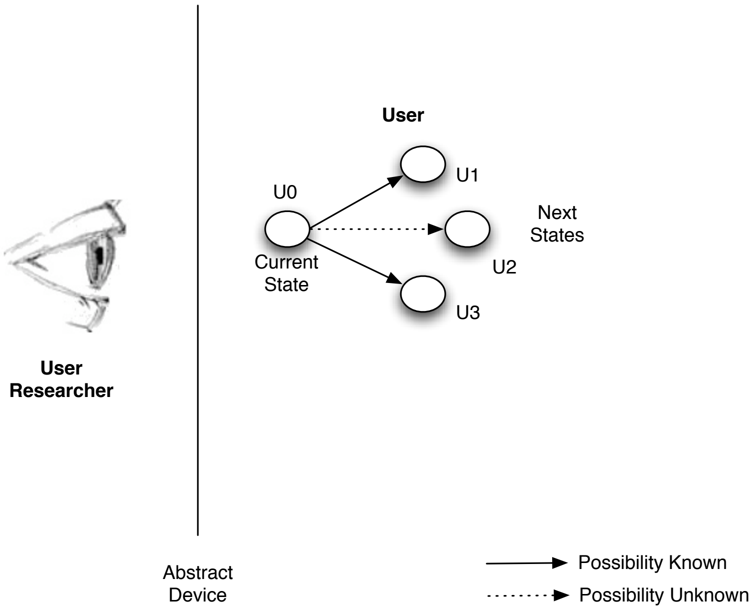


Figure 3.10. Researcher Perspective in User Modelling

The Research Oracle Perspective as a Natural Research Perspective

The research oracle is also of use as an abstract concept, as it is crucial in much IR research. Anytime an IR investigation mentions a user model or makes hypothetical remarks about a search which are not fully knowable in realtime by a researcher during a search the research oracle perspective is being *invoked*. By use of abstractions of search agents and processes in research discourse, one is indirectly creating a search simulation which is not necessarily complete and not necessarily implementable. Thus the concept of higher ostension and the oracles are a particular articulation of what is going on in terms of states and state changes during search and research. There are statements known only through explicit simulation as opposed to 'indirect simulation' through the research oracle perspective referred to above, which would be infeasible for the traditional researcher R_{trad} to know through traditional investigative IR research; perhaps such statements can only be deduced through a large number of computations. Taking the traditional researcher's knowledge as the set of knowledge statements $K\left(R_{trad}\right)$[17], and similarly for the researcher using simulation $K\left(R_{sim}\right)$, then $\exists S\left(S \in K\left(R_{sim}\right), S \notin K\left(R_{trad}\right)\right)$. A complete understanding of the relationships between $K\left(R_{trad}\right)$ and $K\left(R_{sim}\right)$ would require extensive study, this is especially since the user agent contributes significantly to $K\left(R_{trad}\right)$ yet when simulated it only adds some simple user modelling concepts to $K\left(R_{sim}\right)$[18]. Further, the user researcher's knowledge $K\left(R_{user}\right)$ maybe shared with $K\left(R_{sim}\right)$ to make more realistic models, hence the need for an involved study.

3.5 User's Self-Knowledge and Simulation

The prior sections discussed particulars about search states and their changes, this section focuses on reasons for changes of search states. In the traditional model a system exists to serve the user in 'fulfilling their information need', in the ostensive model it is assumed that every user interaction indicates something about the information need (IN) then during the course of a search the states change due to the successive understandings of the user and/or system about the user's IN. What then is IN? Traditionally it refers to knowledge/statements 'missing' in the user, however does the user consciously wish to fill a set of gaps in their knowledge? Consider the following characterisations of user states:

---

[17] $K\left(X\right)$ is further elaborated in Chapter 5, at this stage it is a function that outputs all statements known by agent X.

[18] This depends on the complexity of the model as it is possible for particular types of user model to generate unexpected and/or randomised behaviour potentially adding much more to $K\left(R_{sim}\right)$.

1. User does not know that he/she does not know (compound ignorance)
2. User does not know what he/she wants to know (contextual ignorance)
3. User knows of some things he/she does not know[19]
4. User wants to know something that he/she does not know
5. User does not know if he/she wants to know something that he does not know
6. User does not know what he/she wants to know out of the things he knows he does not know

In the above cases if the user knows something and wants to *refresh* this knowledge then this is equivalent to (3) since it can be said that the user does not know the experience of this refreshing until it happens. Hence the phrase 'refreshing knowledge' indicates that prior to the process of refreshing (each process being unique[20]) the process has not been experienced by the user thus the experience is itself a gap in the user's knowledge. Statements (3) and (4) together describe a typical internet browser whereas (6) can refer to a researcher trying to find a field of investigation. Characterising the user's knowledge and wants in such a way is invoking the research oracle perspective as one is assuming to know statements or lack of statements in the user's mind. The problem is that the user him/herself may not be able to tell us if any of 1-6 is true, nor can a researcher deem any of the user's indications as true utterances of their state. The traditional researcher cannot objectively analyse the user's mind to deduce where the knowledge gap lies but more fundamentally what is a knowledge gap and how is it ideally alleviated, is it by providing the user certain documents and/or experiences or giving a partial experience and a particular amount of time to let them fill the knowledge gap by self-deduction? Any resolution to these questions is context specific, instead these complications can be removed by simply considering the simulation approach in which interpretations of the above characterisations could be implemented in terms of a user model in order to facilitate testing of hypotheses like 'do users with states like (3) and (4) search longer than those meeting (6)?'. In order

---

[19] In this case the user knows how to *refer* to their missing knowledge without knowing the details of it. For example, they may know that they do not know the capital of a country.

[20] This assumption is based on the argument that as each moment of time is uniquely occurring then the state of mind in that time period is unique so any process of state change is necessarily contingent.

to address any user based issue, especially that of defining information need rigorously one is forced to adopt the research oracle’s perspective (in the direct sense) of search which assumes user simulation thereby moving the investigation out of the realms of traditional search.

A model of user’s knowledge can be a set of statements containing all except one statement from the set of all user-knowable statements and take the search objective to be to complete the set. A user that knows that they are missing a statement which they are trying to obtain through search can be represented by a set $U = (s_1, ..., s_{n-1})$ where there would be a sentence $s_1 =$‘$\ldots n - |U| \cdots$’ is a statement referring to the size of the set in which it exists, thus the truth value of s_1 can be used to decide if the search is complete. Modelling a user’s self-knowledge is central to creating realistic user simulation[21]. Dealing with simulated users is similar to dealing with traditional expert systems where the user is the expert in himself/herself and means that many of the advantages of expert systems also manifest themselves with search simulation. One such advantage is that simulated users can be copied and shared among investigations leading to a level of consistency in research that could not be traditionally achieved even if the same human users were to be shared among different searches.

3.5.1 User Simulation

This section addresses the issue of making user models for simulation. First some characteristics of the user’s knowledge are addressed followed by a discussion of user modelling techniques. The function of a user agent is to pick out the next state from the system agent’s set of possible future states according to the user perspective as shown in Fig. 3.6. The user agent picks the next system state by processing two pieces of information, the static or ‘pre-search’ knowledge and the ‘dynamic knowledge’ that it has gathered during the course of the search process both of which are then re-interpreted in light of the choices it now has for new system states. The way in which the static and dynamic knowledges are interpreted, combined and the method of deciding the next state all depends in the

---

[21] User modelling is briefly addressed in Chapter 6. As the user is a type of system it is addressed (by symmetry) in addressing the system. The justification of user models through real user based research is only indirectly addressed in this thesis.

case of a real user on their character, personality or other human traits and therefore these human traits are precisely the customizable variables for simulated users. For a real user to pick a state it needs to know how to interact with a system, thus there is knowledge about using the interface shared by both user and system prior to the first user interaction. Apart from the basic interaction knowledge the user learns implicitly about the workings of the system. For example, when using Google search the user is able to eventually deduce that the system is using keyword matching to return results, this evolving knowledge about the system, or dynamic knowledge, influences the user's further interactions. What is it in the user's knowledge precisely that changes and how?

Given the generic conceptualisation of agents and their knowledges in this chapter, the changing dynamic knowledge corresponds to a changing set of arrows in Fig. 3.2 and also to refinements of the statements within agents. For example, during the start of a Google based search a statement in the user agent referring to their understanding of the system can be 'this reads what I write and gives me a list' which is an abstraction (as indicated in Fig. 3.1) of the mental state of a real user. During the course of the search this statement could be refined to 'this thing reads what i write and gives me a list with documents which have my words in them'. For the purposes of creating a user model of the user referred to in the prior example these statements must be transformed into forms that can be operationally employed by a precise decision function which the simulated user will use to pick the next system state. If the knowledge of a simulated user requires to be represented by precise statements then so does that of the system, as one need not only conduct theoretical investigations but also be able to run computer simulations to verify that which cannot be verified by 'paper-investigations'[22] alone. What is then to be the language for expressing *statements about what an agent knows* and what the corresponding operational framework that shows how statements are used and how they change? This question is addressed by the following chapters which establish such a framework.

---

[22] This refers to verifications of hypotheses that require repeated processing by computational methods, as opposed to manual deduction.

Generating User Models through Analysis of Systems

In addition to requiring a framework one needs to address how a new user model is made and how real users are to be employed (if at all)? First note that a new user model can be generated from a system model. As if the user is the entity that picks new system states then given the set of possible future states each choice for the new state corresponds to a different user model. Although user models can be created in this way by low-level exhaustive modelling using possible system states it is not feasible if it were the only solution, as for example in the Google system, this would correspond to as many user models as there are queries in all words and their combinations multiplied by all possible browsing patterns beginning with a document on the search results page. Rather what is needed is a smaller set of user models that represent a user or user group to whom the search investigation pertains or a set of models representing an average user population if the investigation pertains to all user types.

In a typical modelling process a set of example user models would be generated by the system designers that are representative of the functionality of the search system. Given a description of the type of users to be investigated the user modeller's task is to deduce a representative set of user models. For a particular real user it could be the case that a combination of user models employed strategically through a search leads to a more realistic representation of the real user than any single model. Thus the structure of the investigation as depicted in Fig. 3.11 is such that user modellers and search researchers could work independently much of the time. The instance when serious collaboration is expected is when the generated (from system) user models fail to capture real user behaviour to any satisfiable degree as then the two parties need to decide why their system fails with respect to real user; such collaboration would almost certainly indicate that the system is potentially a poor one. The realism of a simulation investigation is bounded by the realism of the user models it employs, to measure realism the statistical methods employed in traditional IR user-studies are required. Apart from measurement of realism real users need to be studied from the cognitive science perspective to create user models with complex decision functions for picking future system states and in turn advanced system models that could potentially generate such involved user models. Other than

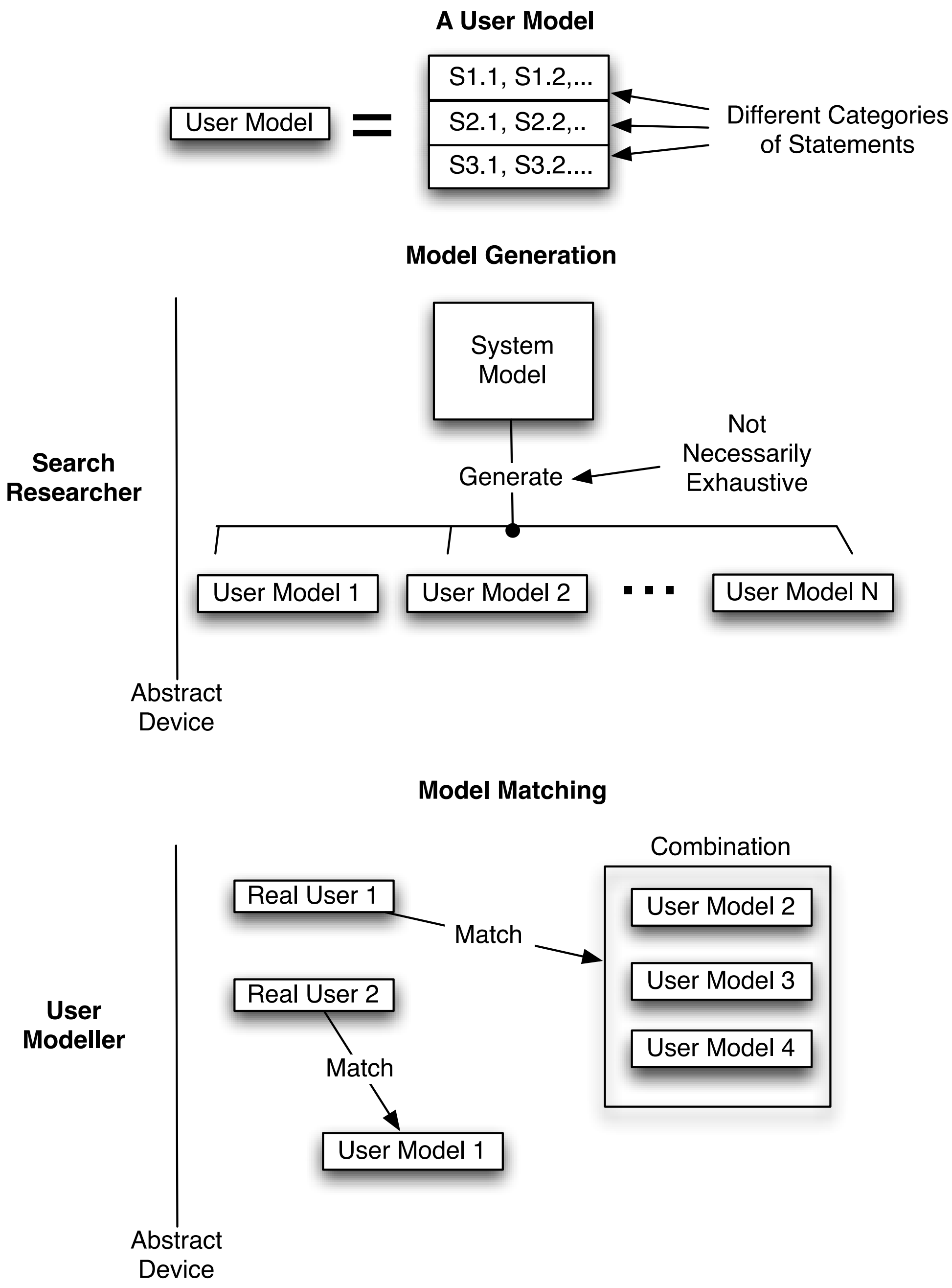


Figure 3.11. User Modelling

this the real user is outside simulated investigations which entirely become a technical, computational discourse.

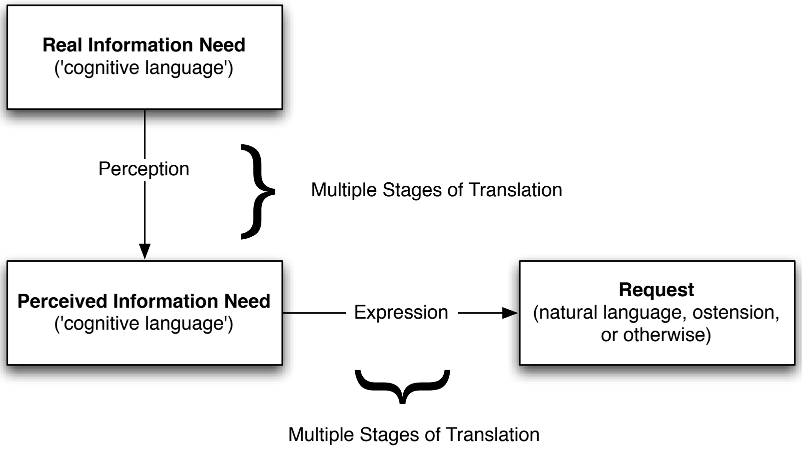


Figure 3.12. Information Need Path

3.6 Non-Determinism and High-Level Search

In the traditional system and researcher perspectives the state selection by the user denotes a non-deterministic choice. The state the system will be in next is not deducible by most users as that would require them to know both the workings of the system and the context of documents in the collection, so from the user perspective future system states occur by non-deterministic choice. In the traditional user perspective the user's information need is assumed to be changing as it evolves from abstract cognitive activity (real information need) towards conscious thought (perceived need) and finally as a user interaction as in Fig. 3.12. Between these stages are perhaps many other steps of deduction and/or understanding[23] of the need. For the user whose need is yet to surface to his/her consciousness, when it does surface is then a non-deterministic choice since the user (1) did not consciously know the need and (2) perhaps did not have the potential of determining until it just happened[24]? Finally, the state being picked by the external oracle is a non-deterministic choice from the research oracle's perspective. In order to make realistic user models, but more generally, for formal discourse into state-based models of search, it is useful to consider such issues as the concept of non-determinism, which has been implicitly related to ostension in prior discussion and is a key issue for computability and complexity investigations that can potentially apply to research on simulations. The concept of the higher ostension for traditional search is that in the perspective of the research oracle

---

[23] Internal process of interpretation, or in general, cognitive activity.

[24] This is a philosophical issue not directly addressed in this thesis.

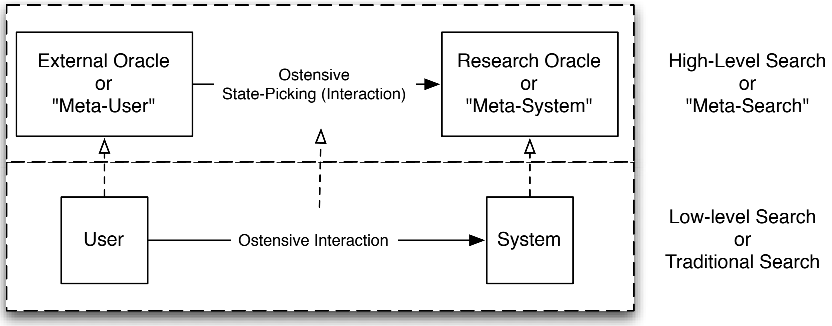


Figure 3.13. Summary of Relationships between Search and High-Level Search

the external oracle causes the overall search state to change by picking the next state out of a set of possible future states. In this respect the research oracle can be called the *meta-system* whereas the external oracle is the *meta user*. The meta-user ostensively specifies the next state meaning that according to the meta-system change 'just happens' almost as if *the meta-user pointed at one of the states expressed by the meta-system's perspective*. On the (lower) level the user ostensively selects the next document on the ostensive interface which corresponds to the next system state. Thus, it appears that the interaction on the level of oracles resembles a search on the level of users and systems, this is illustrated in Fig. 3.13.

Along with the WWOS concept from the previous chapter (Section 2.5) where a very implicit conceptualisation of search is established, the observations in this chapter suggest that perhaps a new definition of search, and of information need, the satisfaction of which is the purpose of search, is required to accommodate the conclusions of prior discussions. This is addressed in the following section.

3.7 What is Search? A Dialectic from First Principles

In the previous section the discussion started with traditional IR toward a state based abstraction of the search process to understand the agents and their interactions during a search. This section approaches the problem by asking fundamental questions about search which are abstract and lead to a concrete definition of search accommodating the issues previously discussed that the traditional conceptualisation of search is unable to address, thus it addresses the conceptual problem. This then acts as a precursor of Chapter 5 which attempts to build from this concrete description an operational framework for search investigation or 're-search'. The discussion in this section is of the dialectic form[25] and is intentionally blind to that of the previous section until the end this is so that these two methods can be shown to approach the same conclusions thereby providing a stronger (mutual) justification for the conceptualisations provided in the following chapters. Initially the questions address the purpose and goals of search leading to discussion on the nature of search agents and ending with a resolution of the conceptual problem of the previous chapter by means of an alternative definition of search.

Q. 1. *[Purpose] What is the purpose of retrieval?*

According to traditional IR it is to fulfil information need but then IN needs to be defined. Instead, consider a different approach by observing that the purpose of retrieval is necessarily embedded in the answer to question Q. 2, as the ideal retrieval agent which is synonymous with a system agent is only ideal as it fulfils its purpose in the best way. Hence one can say that *Q. 2 ostensively answers question Q. 1*. The notation is 'retrieval agent' instead of 'system agent' to direct the reader away from any concrete notions of a system and towards an abstraction[26] of it.

Q. 2. *[Ideal] What is the ideal retrieval agent?*

The agent in question Q. 2 is the **abstract retrieval agent** which shall be denoted RA and it can be the solution to a retrieval problem being of artificial (physical) form such as a computer program running on hardware like traditional search software, a human being (such as a teacher or librarian) or an abstract form. Hence the agent is not restricted to being something that necessarily exists and does not require to be physically realizable. This question is addressed in two ways, the scenario and ontological approach.

---

[25] As this discussion is itself 'search' then the best way to talk about it is by ostension and in written form the question-answer method is closest to the 'interact-react' or ostensive definition method of search; hence the dialectic form seemed the clearest way to express the fundamental aspects.

[26] This need not necessarily be implementable.

3.7.1 Scenario Approach

The first technique is to describe without any formal definitions, in subjective terminology, an ideal retrieval system. Hypothetically and in informal terms it is proposed that the ideal retrieval agent is a type of oracle which detects a user's question with minimal effort required by the user to express the question, and then provides an answer that maximally satisfies the user. The goal would then be to define, further investigate and implement such an oracle or an approximation of it and create evaluation methods which would test the effectiveness of the oracle according to devised measures of user satisfaction. A research structure would emerge held up by the principle of least effort from the user side and maximal satisfaction from the system side[27]. A problem in starting with a subjective proposition is that one now falls into problems trying to describe these very subjective terms, as what is meant by a *question*, and how is it indicated? Any research strategy rooting from an initial subjective description of an ideal system will need to consider how the user interacts with such a system, assuming that user and system are separate entities. It could be the case that the distinction between an user and a system is blurred. For example, an expert system can take the place of a librarian in suggesting directions for investigations by user's interested in specialist topics. If the expert system uses a retrieval system in turn, then it is a user in that context while in the perspective of a user the expert system is itself the system agent. In the interests of theoretical research it is useful to consider the expert system and user as one entity from the perspective of a system agent and similarly it could be useful for analysis from a user perspective could to see the expert system and underlying retrieval system as one. In the perspective of WWOS (and Web 2.0) as discussed in Section 2.5, the system is potentially the operating system and any software running on it; therefore, to accommodate these factors it is proposed that *the distinction between user and system should not necessarily be strict*; separation is addressed in Section 3.7.3.

---

[27] Least effort and satisfiability are both to be judged by a user in traditional search, however, it is assumed that they are both, in the case of simulated search, well defined concepts. However, the least effort measure directly ties to a measure of realism of the simulated user model, while satisfaction is not seen as such, thus the latter is assumed to be "more objective" and hence a system property.

3.7.1.1 Ontological Approach

The second technique of answering question Q. 2 is to ignore any initial vision about an ideal system involved in a scenario, and instead try to define rigorously the terms: ideal and agent. In the proposed question, it is the retrieval agent that aims to be called ideal and any definition of ideal will be required to address a notion of user satisfaction, which means the problems of the previous approach exist here (as ideal and agent are subjective terms). An additional problem is that there is no ‘big picture’ (or context) in this approach which may help to define these terms. In the previous approach, one goal for the ideal agent or the oracle was for it to require minimal user effort to work. There are no such predetermined goals here, and so there is more to define. Thus, there is a dilemma, if one starts with the scenario approach then the the constituent parts of the description need to be defined thereby leading to the ontological approach. Instead, if the ontological approach is taken then a context is required so one is lead to the scenario approach. These two approaches are expressed here as IR research can be classified as one or the other.

Scenario Approach vs Ontological Approach in Practice In traditional IR the scenario approach refers to investigations assuming IN satisfaction equates to agreement with relevance judgements and minimal user effort equates to satisfactory answers on a questionnaire. The ontological approach in IR refers to making models that formally define relevance in terms of mathematical constructs which they test by comparison with relevance judgements thereby linking back to the scenario approach. Information seeking investigations, and investigations inclining towards user analysis are scenario based approaches whereas much of the IR research as distinguished in [IJ05] is technical and ontological in approach. Traditional IR research does not usually make this distinction; the reason for separating them in this way is to make the point that perhaps they should be separated so that technical IR research is allowed to be independent from scenario based research, however as the satisfaction of relevance judgements is assumed to be required for technical research it is proposed that the ontological approach is unnecessarily dependent on relevance judgements and hence on the scenario approach. The criticism here is toward the assumed status of relevance judgements as the definitive criterion for deciding the quality of technical IR models or quality of any ontologically approached research. It is the case that researchers do not see the status of relevance judgements as a definitive criterion for quality but instead as the best criteria they currently have as there are not many alternative measurements that are useful for classification of the quality of research.

Validity of Evaluation Results over Time If a set of relevant judgements are used to test different systems and later found not to adequately represent the greater corpora and user base which the judgements were approximating, then the results would not be adequate. As the user base and corpora, for example of the internet, are changing daily including the sub-nets like the blogsphere which admit different search requirements such that a search engine effective for searching the internet on the whole may be poor for blogs and vice versa. Thus an IR experiment done with a system model some years ago may not be valid for the corpora of today.

Classes of Search Problems and Models Ideally, one would presume that given a new search problem one could identify all prior search models and be able to deem its suitability to the new problem without exhaustively testing them using new relevance judgements or user studies. In the case of algorithms analysis [Pap94, GJ79] new problems are reduced to prior problems by way of classification of associated algorithms into complexity and/or computability classes. For IR, this would correspond to creating alternative measures that are used to form a set of classes[28], independent of subjective relevance judgements, that can be used to structure IR research; such that for a future problem the main task is to formally reduce it to a prior problem for which a system has already been designed. If a reduction is not possible there will be insights into the reason for this which itself would be useful. For example, a search problem involving agents, an interface and a user's interactive behaviours could be classified according to pre-defined measures such as search complexity[29] or satisfiability[30]. As for the traditional approach that always depends on relevance judgements, one could overlook a good quality system model whose quality is perhaps only revealed through a future application, thus it could lay in the research archives

---

[28] In the way a measure on a set can partition it into subsets according the measured value of the items.

[29] A measure denoting the overall complexity of user interactions, search goals and search system, or one could envisage a family of measures pertaining to the complexity in each of the components.

[30] Denoting the extent to which the search in this search problem meets its goals, or in the traditional sense, the extent to which the user's IN is satisfied.

failing to show up through current methods of 're-search' of search. Ironically this failure is due to poor (re) search in the first place, but it is difficult to do better search research with dependency on relevance judgements[31].

3.7.1.2 Dilemma due to Relevance Judgements

The need for relevance judgements is a dilemma for the above techniques. Instead, if a set of alternative measures are created that are independent of the issues of subjectiveness surrounding the relevance judgements extracted from user studies then the problem subsides. If for example there are a set of approximate relevance judgements that can be generated by algorithms that estimate user behaviour, then given that the nature of these judgements and techniques of generation would be well defined one is then able to formally classify system models. As improved relevance judgement approximating algorithms are created for different types of users the previous systems can be easily compared to newer systems as the prior approximation algorithms can be formally compared to the new ones. The main issue then becomes one of finding good formally defined approximations to a real user or a real group of users, so that an *approximate user* can be formulated that formally expresses what it means for it to be satisfied as required for the scenario approach, and thereby also suggesting what is idealness as per the ontological approach. Inherent in the question of the ideal system, is the idea that real users or human beings must always be the judge. It is true that judgments of real users are of great concern when answering questions about the effectiveness of a system in satisfying them. However, it is difficult to deduce definitive answers based just on user opinions, as users will differ in opinion or the same user may differ at different times. Statistical methods are used to tame the ambiguity over user opinions but ideally user judgments should be more definitive so one can revert to statistical methods as a last resort. Approximate users allow research to be classified according to alternative measures while revealing their expected level of error so that the realism of the result is itself another way to classify research. The idea is that as research into user approximations that runs in parallel to system modelling generates more realistic users the quality of the prior systems are automatically re-evaluated using simulated eval-

---

[31] Thus doing good research is related to doing good search (and vice versa), and writing about search research, due to the conceptual self-reference, is limited by that which is aimed to be improved i.e. search.

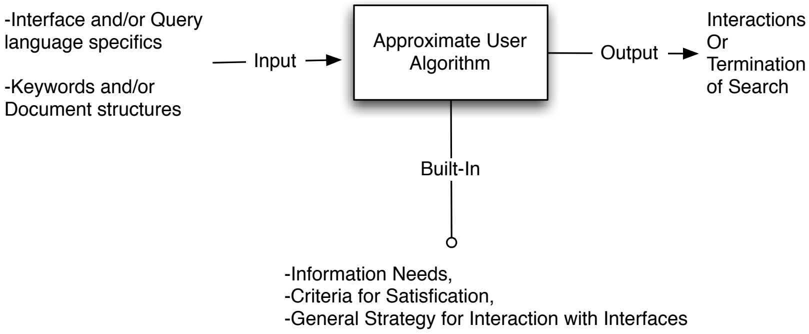


Figure 3.14. Model for an Approximate User

uation now that both the user and the system can be expressed as a set of algorithms. However for this to be anything more than conceptual thinking there requires to be an operational framework in which approximate users can be created; there also needs to be a way to sort out the fact that an user definition is tied to a system interface as it needs to know how to act. However, how would a user algorithm adapt to new interfaces? This depends on the particulars of the formalisation, and is briefly addressed in Chapter 5, but a black-box diagram of the user approximation algorithm is given in Fig. 3.14.

3.7.2 Approximate Users

In creating approximate users, in order to make any opinions from the user into definitive ones forming part of the behaviour of the program approximating the user, it is required to understand other factors affecting the user judgement so that the logic behind the opinion is clear. Deducing all these other factors is non-trivial if not impossible, especially as the user may not know or be able to express all the reasons why they are making the judgement. A strategy for research which temporarily avoids the confusion of user opinion is that of initially ignoring the real user and their opinions all together, and replacing them with a general entity. This general entity known as an **abstract user** denoted `AU`, as in the case of the retrieval system agent, is either a human, a computer program (an approximate or simulated user) or an abstractly specified form (mathematical model, not necessarily implementable, or a description of a cognitive form). All approximate users are a type of abstract user whereas not all `AU`'s are approximate users as some `AU`'s are constructs for theoretical investigation only. The idea is that there will be no confusion as to the opinions of these abstract users, as they are precisely defined; and the realism of a particular abstraction is left to an alternative investigation[32] since the quantification of the extent to which a human user resembles a corresponding approximate user would require statistics[33]. The traditional use of statistics to deal with ambiguity of relevance judgements is still existent but it has shifted, so that it is instead to be employed in the formulation of measures quantifying the realism (or to quantify error due to non-realism) of an approximate user model. The advantage of this shift is that a set of definitive results can be obtained through research about the effectiveness of retrieval agents on approximate (and also more abstract) users while the approximation error is a separate issue which can be tackled in a separate investigation. A further key point that requires to be made explicit at this point is that the most general relation between the `AU` and the `RA` in addition to the trivial relation of them being objects, is that they are engaged in a process and are *separate* objects. The following discussion outlines some conditions for agents that precede their formalization in the following chapters.

3.7.3 Separation and Association of User and System

Any `AU` is a separate entity from a `RA` with respect to some process Π in which they participate. This requires there to always be a way to distinguish between these entities in the framework in which they are formally specified. There also exists some process Π which associates any `AU` with an arbitrary `RA`, so that one can take any abstract user and set it to interact with a system agent even if other intermediary agents are required for compatibility[34] purposes. Association is required to indicate that one cannot make an abstract agent that is completely unrelated to another. The process Π is more rigorously defined in Chapter 5. The notion of separation may raise the question of if it makes sense to create a Π in which the `AU` is defined exactly in the same way as a `RA`, such that it is an interaction between two systems or a game between two players with the same strategy[35].

---

[32] Unlike in traditional IR, user studies can be done by a researcher separately from search engine evaluation, thus there is a change in research method.

[33] Assuming socio-psychological studies would be required to find out about real user behaviour.

[34] Perhaps a user model is designed such that it is restricted to interaction in a particular way that prevents it from *knowing* how to use a particular system, thus, a 'middle-system' or proxy system is required to translate in between the user and system. The realistic example is where the system is a highly technical machine used by a librarian acting as proxy between the user and it.

[35] As addressed formally by game theory [Gin00].

A particular case of this can be seen in the case `AU` is a real user and `RA` is also the same user where the participation between the human and himself/herself is that of them talking to the mirror, or self-analysis. To resolve this issue, there requires to be some concept of role that needs to be clearly defined for the agents involved in the process specifically to state which agent starts the process. With regards to prior discussion question Q. 2 can be revised for clarity in light of the notion of the abstract user to the following question.

Q. 3. *[Relative Idealness] What is the ideal retrieval agent* `RA` *according to the abstract user* `AU`*?*

The revised question is now implicitly directed at the researcher who is assumed to know `AU` and `RA` in definite terms. This revision of Q. 2 was intended to clarify that judgments about an ideal `RA` comes from a researcher and not directly from a `AU`. A researcher can use a human user to help them answer this question by deducing the realism of a `AU` (among other factors) but they need not, and instead can conduct an investigation with other measures for idealness. A researcher is defined as a role which investigates the above question. Question Q. 3 is the archetypal research question for usual search investigations, but if taken in an absolute sense it refers to a theoretical ‘maximum’ or ultimate retrieval system. The Memex vision of Vannevar Bush [Buc92] is conceptually one such answer for the prior unrevised question, but it is of the form of the scenario approach, in that it describes a set of scenes in which a human user works with a retrieval agent but does not strictly define the agents therein or concepts of satisfaction. There can be no absolute ideal `RA` but an ideal `RA` with respect to one or more `AU`’s. As concern about real human users is separated from analysis of ideal `RA`’s, the researcher is given more ‘administrative space’ to investigate a wider range of questions regarding a retrieval agent without being concerned about the physical realisation of their results. The premise for such an attitude to research is that there will eventually be methods to validate or invalidate the specification of an abstract user using real human user data but the bottleneck of having to simultaneously prove correctness according to subjective measures as well as make systems, is removed, thereby allowing the theoretical research proceed in the meantime. This is comparable to scientific research in the area of algorithms and complexity where theoretical investigation without the need for direct application to real-world problems is a necessity[36]. It maybe the case that even if the abstract user definitions are found to be invalid or not physically relevant, the research based on it is important or useful in some other way. Some example research questions that are easy to ask in the realm of abstract users and systems, but difficult to approach in traditional research, are presented below.

1. What is the ideal `RA` for an `AU` approximating a blind human user?
2. Which `AU`'s will never be able to effectively use some particular set of `RA`'s?
3. Is the effectiveness of a `RA` for some exhaustive (complete) set of `AU`'s always limited unless they learn to *use* a particular feature of a retrieval agent?

In research question one an absolute result is desired for modelling a blind user which means that there requires to be an approximate user model (with the least error in realism) augmented with limiting values denoting the lowest *possible* error values for any approximation of blind users. There could be a class of such user models $\mathtt{AU}_B$ with $\alpha_\Omega(B), \alpha_\Theta(B)$ denoting the upper and average case approximation error of the models respectively[37]. In fact, since each user model is essentially an algorithm techniques similar to those employed in the research of complexity of algorithms, could be used as a basis for classification of these models where instead of run-time and storage space required for an algorithm there is an additional set of parameters one of which is, approximation error. Instead of finding the ideal `RA` for each blind user it could be more convenient to find an ideal one for the average class $\alpha_\Theta(B)$ and since it is known to be the average, the found system (`RA`) could be said to be ideal in the case of the average blind user given that $\mathtt{AU}_B$ is representative of the real user population[38].

Problem two includes the temporal operator 'will', the existential operator 'never' and the 'effective use' measure. To address this problem one needs to find all ways of associating a

---

[36] As it deals with abstract relationships between algorithms which can be applied to real-world problems, so much of the research is *potentially* useful to practical applications and not necessarily of immediate benefit. Compare this to the common research method in IR following the laboratory model which requires particular results pertaining to measures of effectiveness of a system.

[37] In practice these classes would be derived through experimentation with real users augmented with statistical or other methods to generate the error in a particular user model; this is not further addressed in this thesis.

[38] Which would be discerned through user-studies and statistics.

set of user models, say $\texttt{AU}_E$, to the particular set system models, say $\texttt{RA}_E$, which would be done through a set of processes Π_E. The process along with the agents are supposed to expose the ways of interaction between the agents and a stopping condition to say when the process terminates. An effective use measure would observe the interactions and assign each particular process a numerical value which if it is below a threshold would admit the agents in that particular process into the answer set of the problem. Problem two is an archetype of the kind of investigation that could aid in the design of systems for a particular set of users or a particular user type (e.g. 'the common user' or 'the expert user').

Problem three concerns the effectiveness of a system model for all user models where the event of a user, having learned to use a system feature, would be indicated by changing their style of interaction between the user and the system in processes in which they associate. A similar set of processes for all users would take place. An exhaustive set of user models may not be possible thus one can take representatives from different user groups/types and seek a probabilistic answer, for example that $0 \leq \mathbf{Pr}($Effectiveness of $\texttt{RA} \leq e$ given $\texttt{AU}$ "Learned" a feature of $\texttt{RA}) \leq p$. Research of this type could be very useful for deciding whether to train users to use a particular feature due to its effectiveness (if the above probability is large enough or to remove the feature).

Being able to think of a system and its abilities by putting the real user in the background, and only thinking of the simulator is convenient as one is not dependent on the presence of physical users but can use them at any suitable stage in the investigation. So far it has been identified that without a strict definition of a user, any approach taken to answer question Q. 2 leads to problems. A revised version of question Q. 3 explicitly demands an user definition, and thus guides research by making clear the inherent complication in addressing the original question. Further, it was proposed that there are some very general relations between the agents which exist since the agents participate in a process and therefore need to know things about one another to interact. This process is a communication process, such that there is an exchange of information during interaction between the agents. One could assume at this point that Π denotes agent communication of a very specific sort and is not arbitrary. However, in light of the previous section and the discussion above, the concept of search is re-defined so that Π may indeed include an arbitrary communication process. As in the case of a search process there are retrieval specific goals and measures as suggested in the scenario/ontological approaches, what if any, are the goals of an arbitrary process? Is there a general goal? Can this arbitrary process be seen as a search process, and is it of use to perceive it as a search process? The next section addresses some of these questions.

3.7.4 Resolving the Conceptual Problem

This section directly addresses the conceptual problem in Chapter 2 by following from the above discussion, recall that all other IR problems depend on it (as articulated in Section 2.6). Initially some observations are made which further elaborate the conceptual problem a resolution of the problem is then presented. First, it is important to realize that relevance feedback can be implicit: any interaction with a search system is a relevance feedback. This is equivalent to saying that *the user ostensively picks the next system state* as elaborated earlier in Section 3.1. Second, a search interface is not necessarily explicit, and can be defined as an interface by which an user's cognition is influenced, which thereby includes any interface that can be sensed assuming any sensorially influenced activity modifies cognition in some way. As exemplified by information seeking, a user does not require a computerized search engine to search, any interaction with another human or agent is also search. Hence it is proposed that any communication process can be viewed as a search[39]. The higher ostension principle applies to to the interaction between the research oracle agent which is in a particular state and the external oracle that causes these states to change. Given any process with two or more agents the research oracle for that process would be the hypothetical entity to whom all possible future states of the process (and hence the agents) are visible and the external oracle is that which ostensively picks the next states thereby inducing state change.

All Processes are Search Processes Following from this it is proposed that any process in which *change* occurs can be represented by the oracles interacting by higher-ostension so every process is a 'high-level' search process (as mentioned in earlier discussion). Gener-

---

[39] As search has no specific definition (that could restrict scope) it is defined with maximum scope as it is useful to do so for modelling and for addressing the definition problem. This implies that a method of modelling search as defined here is a type of "process modelling language".

alising and redefining further, it is said that search **is** this notion of the high-level search[40]. This addresses the conceptual problem of 'what is search'.

General, Non-communication processes It is assumed that every physically existing process has change occurring in it, however insignificant, until it ceases to exist. It is also assumed that an abstract process, one not directly physically existing, exists indirectly, as it requires the existence of physical processes to express it. For example, a mathematical description of a process is an abstract description which can only exist when expressed, either in the mind of a mathematician by physical cognitive processes, or in physical reality. Hence every process, physical or abstract, is a search process by this conceptualisation.

Cognitive Processes The process of one part of cognition influencing the other as in the case of the forming and expressing information need in Fig. 3.12 is also a search as each part of cognition can be seen as an agent causing change in the corresponding part of cognition (which is another 'agent'). It is implied that search is synonymous with change, but then what is a process that changes actually *searching for*? Calling every process search makes the word 'search' a primitive but the original meaning of the verb 'search' is to look for something. The original meaning can be accommodated by saying that any process with change is a process 'searching or looking for change' until it is terminated, thus *state change is the search goal in the most general sense*.

High-Level search and Information Need If information is that which causes representations such as state diagrams to change (as defined in [Mac69]) then every process can potentially accept information as every process can potentially change until terminated. The traditional supposition in IR research is that the purpose of an user interacting with a search engine is for fulfilment of their *information need* (IN). Following on from the discussion above, it is proposed that IN in the most general sense corresponds to a lack of experiential knowledge resulting in an *experiential gap* of which the semantic, knowledge and social gaps are specific examples in order of generality[41]. Assuming that all knowledge

---

[40] Thus, the generalisation does not distinguish between high and low-level (traditional) search but instead defines a search process as a type of process that can refer to low or high-level search.

[41] Whether the relationship between these type of gaps can be formulated as a type/sub-type relationship, set oriented relationship or otherwise is an open research problem and is out of the scope of this thesis.

is experiential, in that even if some knowledge is built into an agent and denoted as a 'fact' the agent is said to be *experiencing the creation of this knowledge within it*, thus as every knowledge has to exist it therefore has to be experienced.

Fulfilling information need means accepting changes ostensively; in the general case it means to accept state change as change is the gaining of experiential knowledge (and the object of search) and *experience is ostensive*. Given any process that associates agents, the process is, by the above discussion, a search, as it changes state; it changes state ostensively in the perspective of the research oracle. Further in the research oracle's perspective the most general IN which is a lack of 'experiential knowledge of the state change', is itself fulfilled. In light of current discussion, an abstract user can never fulfil the IN they are associated as they are defined by a set of physical processes which exist as long they do, meaning that the user is always ready to accept information, to accept change, and therefore always has an IN. This concept of IN applies unambiguously to all agents, with information defined as according to [Mac69] since all agents change over time. Agents have a potential to change therefore have an IN meaning that the IR system has IN according to the generalized definition of IN and not in the traditional sense. All processes change over time, and are therefore search processes. Thus it is proposed that search is not only a process but it is every process, and not only every process but *the process by which any process is known*[42], interpreted, discussed or expressed[43]. Equivalently search is the topic under analysis but it also denotes (at some level) the method of analysis. Defining information and search in this way directly addresses the conceptual problem. The definition problem now has a much larger scope, instead of referring to traditional computer based search it now refers to any process. This allows one to try any pre-existing formalisms for defining processes whether it be QT methods (as in the next chapter) for defining physical processes or otherwise. This does not mean that any search model has to explicitly be able to accommodate the modelling of all processes; as by definition, all search models address state and state change (directly or otherwise) and thus are said to be implicitly modelling a process. Hence, it is the relationship between the high-level concepts of search and IN, and their low-level equivalents, that allow a low-level search

---

[42] See Chapter 5 for discussing of what it means to 'know' in the context of search.

[43] Consider that in the physical world light has a similar special function, it is not only a visible thing but it is that by which all else is made visible to the naked eye

model which is a type (or case) of high-level search that does not have to accommodate all specific low-level cases. The relationship between high-level and low-level search/IN can be perceived in the object-orientated sense in the way a class relates to its sub-class, it can also be seen as type relationship; the representation in Chapter 5 aims to accommodate both these perceptions.

3.7.4.1 Quantum State Collapse and the Higher-Ostension

In a QT model of a physical process, state change is represented as the change of representation of a wave function. Using the concepts in this chapter it can be said that a QT model of a (physical) system is a state/state change model from the QT researcher's perspective of physical reality as illustrated in Fig. 3.15. If the QT researchers model is a closed one that is assumed to know all possible future states of the physical system, then its perspective is said to change to that of the research oracle defined in the context of the principle of higher-ostension. State change occurs by *ostensive definition of the next state by the agent of reality*[44]. The agent of reality here is the external oracle. The external oracle dictates how the change happens while the researcher has no way of certainly knowing the next state. The ostensive selection of the next state in the higher-ostension principle corresponds to the quantum theoretical state change which in the wave-function method of modelling is called a wave-function or state collapse, note that the association *between* possible future states is ignored at this point[45]. Thus there is now a high-level epistemological relationship between QT and our formulation of IR, and indeed between all processes.

It is further proposed that information need is the most general ontology or 'static component' for any process meaning that it is assumed that IN can always be used to give a partial description of the state changing in the process. Information need is also proposed to be the most general purpose of the process such that it can be said that the purpose of any process is to satisfy a (high-level) information need. Any other purposes, such as to retrieve a document containing a query term is said to be a specialised *instance*[46] of

---

[44] The agent of reality is nature.

[45] Whether future possible states are in physical superposition or related by statistical combination of states and the utility of assuming either for modelling retrieval is addressed in the following chapters.

[46] In the way a subclass is an instance of a superclass while being different from it and being similar

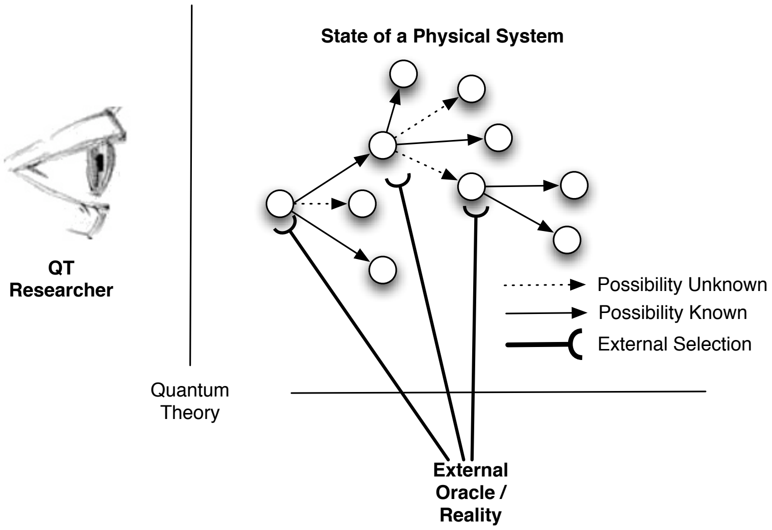


Figure 3.15. QT Researcher's Perspective of Reality

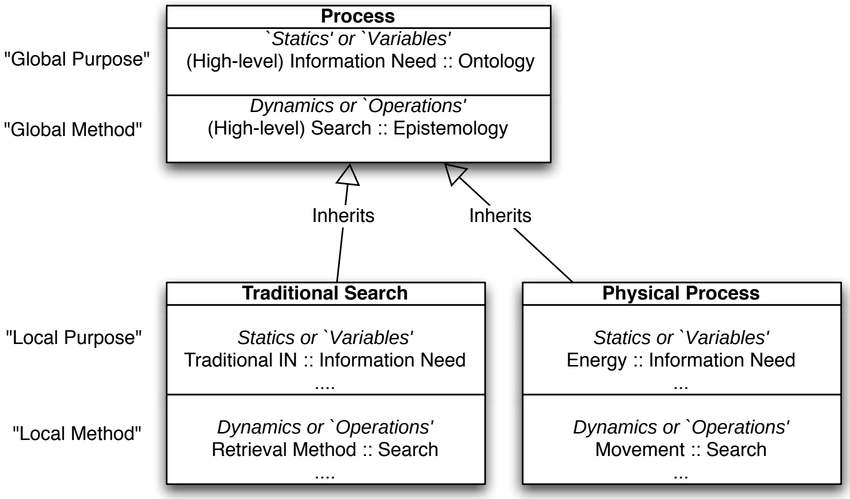


Figure 3.16. An object-oriented perspective of the relationship between search and high-level search where A :: B indicates that concept A is an 'instance' of concept B

satisfying the IN. In a similar way, (high-level) search is proposed to be the most general epistemology[47] and 'dynamic component' for any process, in that, all state changes in pro-

---

due to inheriting properties of it, and in the way the concept of polymorphism can be used to refer to the subclass by the superclass. Thus, all purposes are subclasses of the 'superclass (or global) purpose' of satisfaction of IN and this is why all processes (which necessarily 'inherit their ontology and purpose' from the [high-level] IN superclass) can be referred to by the superclass of (high-level) IN. The representation in Chapter 5 elaborates this relationship whereas Fig. 3.16 depicts it briefly.

[47] This means that (high-level) search denotes the most general method or way of conducting any process or of changing state. It is the superclass for dynamic properties of a process, for any movement, or for anything resembling a change e.g. the function call between components in the matching sub-process of a traditional search would by this definition be of a subclass of (high-level) search as it pertains to the dynamics of the process (since function calling necessarily leads to state change in some component). This is further elaborated by Chapter 5; an illustration is given in Fig. 3.16, see also Fig. 3.13.

cesses can be described by the concept of higher-ostension (including the idea of ostensive picking of next states in the context of observers and oracles) and therefore be regarded as a (high-level) search. Thus, (high-level) IN is a general ontology, static component, purpose and partial description of all states (in a process) whereas (high-level) search is a general epistemology and dynamic component of all processes, a summary illustration is in Fig. 3.16. In simpler terms, search is the way by which anything is known and IN refers to that which can be known and why one may want to know it[48].

3.8 Summary

This chapter showed that all search scenarios can be reduced to a scenario resembling an ostensive search and that this ostensive scenario can itself be generalised so that it resembles on a higher level, a state model. The concepts of researchers, oracles, knowledge and knowables were introduced to characterise how a search process changes according to the perspective of different agents. The principle of higher ostension was proposed to address the way that a search process, which was generalised to any process, is started in the first place. The conceptual problem of Chapter 2 was addressed and with respect to the definition, evaluation and user problems, it was proposed that questions about ideal retrieval systems can only be effectively addressed assuming simulated 'approximate' users. The dichotomy of low-level and high-level IN and search allowed the conceptual problem to be resolved, and also presented a philosophical conceptualisation of search and IN as being a general characteristic of all processes and their purposes respectively. Finally, the principle of higher ostension was related to wave function collapse, a particular phenomena of QT in order to indicate one of the conceptual justifications for employing QT for modelling search. The next chapter further explores how QT could aid search process modelling.

---

[48] More simply, IN is the 'what and why' of a process whereas search is the 'how'.

—Sell your cleverness and buy bewilderment.

Jalaluddin Rumi (1207-1273)

4

On Using Quantum Theory for Search

4.1 Introduction

This chapter investigates the modelling of the state and state change of a search process according to the perspective of the research oracle from the previous chapter. Initially in Section 4.2 the representations of QT in the Hilbert space are presented with details on how the representation is operationalised, then the suitability of these techniques for search process modelling are addressed. Section 4.3.3 considers the conceptual and interpretative foundations for representing physical phenomena using QT's Hilbert space formalism, it discusses the position of observers in a QT and relate this to the concept of observers and oracles discussed in the previous chapter. In Section 4.3.4 I show some techniques of using the Hilbert space of QT to model retrieval concepts from the researcher perspective, in particular the dynamics of a search. I show that the complex field on the Hilbert space can be employed by using the real part of a term weight in a document to represent static information (term frequency and inverse document frequency values) and using the complex part for dynamic information such as probabilities of relevance of the term in the collection. This allows these two types of information to be represented separately in a convenient way. The static and dynamic parts of the weight can then be combined according to a function over the weights of terms and made to correspond to particular semantics[1]. In Section 4.3.4 two simple measures are introduced for combining weights that appear naturally from the representation and corresponding quantitative characterisations of *directional information need change*[2]. In Section 4.3.6, the problem of using typical QT methods to model evolution of a search process is addressed. The complex number representation of weights from Section 4.3.4 is used in Section 4.3.7 for designing the dynamics of a simulation by suggesting the use of group-theoretic representations for the changes in probabilities of relevance of terms. This adds a discrete structure for characterising changes in user interests which is presumptive but is a necessary initial step in modelling simulated users. The applications of QT methods for IR in this chapter indicate that if the semantics of QT could be well understood in terms of retrieval, and retrieval in terms of QT, then

---

[1] According to how one envisages the relationship between static information about a corpus and dynamically changing user INs as represented by changing probabilities of relevance of data items (terms or documents).

[2] For further semantics for IN change refer to the next chapter which relates the semantics of energy from physics and IN to search processes.

both the physical insights from QT and its mathematical framework can be extremely useful for formal search analysis. This leads on to the next chapter which introduces a middle-language between QT, IR and CS which hopes to lay foundations for such 'semantic matching' between these disciplines.

In this chapter, QT is at most partially employed, there are not any explicit quantum effects in information retrieval but instead IR has properties which are conveniently modelled *as if* they represent quantum effects. This thesis employs QT theoretically to analyse IR in the way a programmer employs an application programmers interface (API) for a particular operating system to create software to run on that operating system. The equivalent of the operating system in our case are the mathematical foundations of QT that specifically pertain to properties of the Hilbert space, and the 'API' is our use of it to 'code' the *analyses of IR* each of which corresponds to a 'program'.

4.2 Background

4.2.1 Introduction

This section introduces the required mathematics and its physical interpretation. Initially, the notion of the wave function is presented along with its traditional representation, as a vector in a Hilbert space $\mathcal{H}$. The wave function is a description of the state of a physical system at some point in time. There are two classes of operations performed on the wave function. The *observation* operations[3] are used to extract properties of the physical system. The other type, *evolution* operations, change the wave function to reflect the change within the system over time, space or some other variable. In the high level sense, wave functions, operators, their respective applications and the interpretations of these elements are sufficient for describing any physical system.

---

[3] An observation operation is one which checks if a state has a particular property. Equivalently, one can say that (1) it checks the truth of a proposition which pertains to the particulars of a state or (2) it *queries the state* with a proposition, represented by operators on the Hilbert space, getting back 'answers'/results indicating (with uncertainty) the truth of the proposition. A further understanding of this notion is found in [Gri02].

4.2.2 The Hilbert space and Dirac Notation

This section assumes knowledge of basic linear algebra [Axl99] and aims to only introduce Dirac notation and some mathematical methods for using the Hilbert space $\mathcal{H}$ for QT. A Hilbert space denoted by $\mathcal{H}$, is a vector space with an inner product and norm such that it is a *complete metric space*. In QT, the representation space is a Hilbert space on a complex field, so that a vector's co-ordinates are specified by complex numbers. Traditionally Dirac notation is used to denote vectors and operations on this space, the notation is such that it minimizes the visual complexity of expressing the intended mathematics, simplifying and reducing effort of the overall calculation. A column vector or *ket* is represented as $|x\rangle$ and a row vector or *bra* as $\langle x|$. Computing an inner product between two different vectors require the first to be a column vector and the second a row vector, this is represented, for two vectors $|x\rangle$ and $\langle y|$, as $\langle y|x\rangle$.

A required abstraction is to define the *linear functional*[4], a function corresponding to each $\langle y|$ in $\mathcal{H}$ that assigns to each input ket $|x\rangle$ in $\mathcal{H}$ a complex number. The inner product defined previously can then be re-defined as the functional $\mathcal{I}_{\langle y|}(|x\rangle) = \langle y|x\rangle$. Without loss of consistency and with the intention of simplifying the notation, any vector $\langle x|$ is taken to be equivalent to its corresponding functional as well as being a row vector. Hence, there are two equivalent interpretations of $\langle y|x\rangle$, as the inner product between two vectors and as an application of the linear functional corresponding to $\langle y|$ on $|x\rangle$. This is true assuming the functional obeys the rules for complex inner product spaces. One such rule $\langle x|y\rangle = \overline{\langle y|x\rangle}$(where $\overline{\langle y|x\rangle}$ is also known as the complex conjugate of $\langle x|y\rangle$) indicates that there is a relation between row vector $\langle x|$ and column vector $|x\rangle$, $\langle x| = \overline{|x\rangle}^T = |x\rangle^\dagger$ where the dagger operator $\dagger$ denotes the combined application of the transpose and conjugation operations[5] (order of application of these operators does not change the result). An inner product between two vectors would then require the first to be transformed into a bra (or functional).

As the set of functionals is in one-to-one correspondence with the set of kets obeying

---

[4] A functional is a real-valued function of functions on a vector space, see [Gri02].

[5] Given a matrix with elements $v_{i,j} \in \mathbb{C}$ where each element has form $v_{i,j} = a_{i,j} + \jmath b_{i,j}$ with $a, b \in \mathbb{R}$ and $\jmath = \sqrt{-1}$, the transpose v^T is the map $v_{i,j} \longrightarrow v_{j,i}$ whereas the conjugate is the map $v_{i,j} \longrightarrow \overline{v_{i,j}} = a_{i,j} - \jmath b_{i,j}$.

certain rules, they are said to form a linear space of their own, the *dual* space denoted $\mathcal{H}^\dagger$, so that the definition is $\mathcal{I} : \mathcal{H} \longmapsto \mathcal{H}^\dagger$. Thus, Dirac notation allows functionals to be expressed and used quite 'automatically' in this sense without requiring to directly refer to the matrix representations (which are implied). The sections below summarise the required mathematical concepts.

4.2.2.1 Vectors and Operators

In this section a summary is given of the mathematical concepts prerequisite for studying QT and is adapted from [NC00]. This section also summarises the concept of vectors and functional already introduced in the prior section.

Vectors In summary of the prior section, in Dirac notation a vector or column vector v is written as the ket $|v\rangle$ whereas its conjugate transpose $(|v\rangle)^\dagger$ is written $\langle v|$ which is a row vector such that each entry in $\langle v|$ is the complex conjugate of each entry in $|v\rangle$.

Linear Operator A *linear operator* A is a function that maps $\mathcal{H}$ onto itself, its application is denoted $A\,|x\rangle$. It is useful for analytical purposes to work with the notation $A\,|x\rangle$ instead of the vector resulting in applying the operator. A linear operator between two vector spaces $V, W \subset\mathcal{H}$ is a function $A : V \mapsto W$ with linear inputs, $A(\sum a_i\,|v_i\rangle) = \sum a_i A(|v_i\rangle)$ where according to the conventions of Dirac notation $A(|v_i\rangle)$ is usually written $A\,|v_i\rangle$. The identity operator I_V is an operator defined on a vector space V; $I_V : V \mapsto V$ such that $\forall\,|v\rangle\,, I_V\,|v\rangle = |v\rangle$. The most convenient way to interpret operators is in terms of a matrix representation such that $A\,|v_i\rangle = \sum A_{i,j}\,|v_i\rangle$ where $A_{i,j}$ is the $(i,j)^{\text{th}}$ entry in the matrix A representing the operator A.

Linear Operator Space The set L_v of linear operators on a Hilbert space is itself a vector space as the sum of two linear operators is a linear operator. In addition, zA is a linear operator for complex number z and there is a zero element namely the zero matrix[6].

---

[6] This again indicates that the concept of an operator and that of a matrix are equivalent on the mathematical level in the QT formalism.

4.2.2.2 Inner Products, Eigenvectors and Eigenvalues

Inner Product An inner product takes two input vectors $|a\rangle\,|b\rangle$ and returns a complex number as output, it is usually written $(|a\rangle\,, |b\rangle)$ but denoted in Dirac notation as $\langle a|b\rangle$ and obeys the usual rules of inner products on complex spaces. Any vector space whose vectors satisfy the inner product relation is called an inner product space, a finite dimensional $\mathcal{H}$ is a exactly an inner product space by this definition. The inner product of some $|y\rangle\in\mathcal{H}$ with the transformed ket $A\,|x\rangle$ can be written as $|y\rangle^{\dagger} A\,|x\rangle = \langle y|A|x\rangle$. This can be thought of as the functional $\langle y|\,A$ applied to $|x\rangle$, hence a linear operator in $\mathcal{H}$ induces a linear operator in $\mathcal{H}^{\dagger}$. A simple type of operator known as a *dyad* is defined as $A = |a\rangle\langle b|$, which on application $A\,|x\rangle = |a\rangle\langle b|.\,|x\rangle = |a\rangle\,\langle b|x\rangle$, shows that it does indeed map $\mathcal{H}$ to itself, as $\langle b|x\rangle$ is a constant. The dagger operator can be applied to get a similar result for $\mathcal{H}^{\dagger}$, $(A\,|x\rangle)^{\dagger} = \langle x|\,(|a\rangle\langle b|)^{\dagger} = \langle x|\,|b\rangle\langle a| = \langle x|b\rangle\,\langle a|$ where applying the dagger to the operator is equivalent to applying it to its bra and ket components separately.

The row vector $\langle a|$ is also called the dual of $|a\rangle$ and can be interpreted as a linear operator $\langle a| : V \mapsto \mathbb{C}$. The inner product can be used to define the concept of the length of a $|||v\rangle|| = \sqrt{\langle v|v\rangle}$ and the concept of a *unit vector* u, $|||u\rangle|| = 1$. An orthonormal basis for a space V is a set $[|v\rangle]$ of orthogonal vectors $\langle v_i|v_j\rangle = 0$, which can be linearly combined to represent any vector in V, in addition all vectors are of unit length $|||v_i\rangle|| = 1$. An inner product of two vectors is equal to the vector inner product between two matrix representations of those vectors, provided the representations are written in the same orthonormal basis.

Outer Product If $|v\rangle$ is a vector in an inner product space V and $|w\rangle$ a vector in inner product space W then the outer product denoted $|w\rangle\langle v|$ is a linear operator from V to W whose action is defined by $|w\rangle\langle v|(|v'\rangle) = |w\rangle\,\langle v|v'\rangle = \langle v|v'\rangle\,|w\rangle$. Any linear combinations of outer product operators are also outer product operators. The orthonormal basis for a space V where $\forall\,|v\rangle \in V, |v\rangle = \sum v_i\,|i\rangle$[7] is denoted by $|i\rangle\langle i|$ such that $\sum|i\rangle\langle i| = I$ the identity matrix.

---

[7] Here v_i denotes the i^{th} co-ordinate of vector v.

Eigenvectors In Dirac notation, the eigenvector of a matrix or linear operator A is $|v\rangle$ such that $A\,|v\rangle = \lambda\,|v\rangle$ where λ are the eigenvalues. The diagonal representation for an operator A on a vector space is a representation $A = \sum \lambda_i |i\rangle\langle i|$ where $|i\rangle\langle i|$ are an orthonormal *set of eigenvectors* for A with λ_i being the corresponding eigenvalue. Diagonal representations are also known as *orthonormal decompositions*.

4.2.2.3 Hermitian Operators, Projectors

Hermitian Operators If A is a linear operator on $V \subset \mathcal{H}$ there exists an unique linear operator $A^\dagger$ on V such that for all vectors $|v\rangle, |w\rangle$ in V $(|v\rangle, A\,|w\rangle) = (A^\dagger\,|v\rangle, |w\rangle)$. This linear operator is known as the adjoint or Hermitian conjugate of the operator A. An operator A whose adjoint (its conjugate transpose) is also A (thus $A^\dagger = A$), is known as a *Hermitian* or *self-adjoint operator*.

Projectors If W is a k-dimensional vector subspace of the d-dimensional vector space V, and one constructs[8] the basis $|1\rangle, |2\rangle, \ldots, |d\rangle$[9] for V such that $|1\rangle, |2\rangle, \ldots, |k\rangle$ is an orthonormal basis for W then $P = \sum_{i=1}^{k} |i\rangle\langle i|$ is the projector onto the subspace W[10]. It can be shown that $|v\rangle\langle v|$ is Hermitian for any vector $|v\rangle$ so P is Hermitian, $P = P^\dagger$. The orthogonal complement of P is the operator $Q = I - P$, further, Q is the projector onto the vector space spanned by $|k+1\rangle, \ldots, |d\rangle$.

Spectral Decomposition An operator is normal if $AA^\dagger = A^\dagger A$ which implies that a Hermitian operator is normal. The spectral decomposition theorem says that any normal operator M on a space V is diagonal with respect to some orthonormal basis for V, and conversely, any diagonalizable operator is normal. In terms of an outer product representation this means that M can be written as $M = \sum \lambda_i |i\rangle\langle i|$ where λ_i are the eigenvalues of M, $|i\rangle$ is an orthonormal basis of V each of which is an eigenvector of M with eigenvalue λ_i. In terms of projectors, $M = \sum \lambda_i P_i$ where λ_i are eigenvalues of M and P_i is the projector onto the eigenspace of M corresponding to λ_i, these projectors satisfy the

---

[8] Using the Gram-Schmidt process for example [Axl99].

[9] Note that these kets are labelled by a number instead of a letter; this is a standard enumerative way to label bases. Thus, instead of $|x_i\rangle$ one simply writes $|i\rangle$.

[10] This is independent of the basis chosen for W.

completeness relation $\sum P_i = I$ and orthonormality condition $P_iP_j = \delta_{i,j}P_i$[11]. A normal matrix is Hermitian if and only if it has real eigenvalues.

Positive and Positive Definite A positive operator A is one where $(\left|v\right\rangle, A\left|v\right\rangle)$ is a real non-negative number, if it is strictly above zero for non-zero $\left|v\right\rangle' s$ then A is said to be positive definite.

Unitary Operators A matrix or operator U is said to be unitary if $U^\dagger U = I$ which implies that an operator is unitary if and only if each of its matrix representations is unitary. An unitary operator U satisfied $UU^\dagger = I$ and therefore U is normal and has a spectral decomposition.

Geometrically, unitary operators are important as they preserve inner products between vectors since for any $\left|v\right\rangle, \left|w\right\rangle, (U\left|v\right\rangle, U\left|w\right\rangle) = \left\langle v\middle|U^\dagger U\middle|w\right\rangle = \langle v|I|w\rangle = \langle v|w\rangle$ which suggests that for an orthonormal basis $[|v_i\rangle]$, $|w_i\rangle = U\left|v_i\right\rangle$ is also an orthonormal basis as U's preserve inner products. For any two orthonormal basis $[|v_i\rangle], [|w_i\rangle]$ the outer product sum $\sum_i |w_i\rangle\langle v_i|$, is an unitary operator.

4.2.2.4 Operator Functions

Given a function $f : \mathbb{C} \mapsto \mathbb{C}$ it is possible to define a corresponding matrix function on normal matrices (or a subclass such as Hermitian matrices) by the following construction. Let $A = \sum_a a|a\rangle\langle a|$ be a spectral decomposition for a normal operator A, define $f(A) = \sum_a f(a)|a\rangle\langle a|$[12]. This process can be used to define the square root of a positive operator, the logarithm of a positive-definite operator and the exponential of a normal operator.

Trace The trace $\mathrm{tr}(A)$ is a function over operators that returns the sum of the diagonal entries of the matrix representation of $A \sum_i A_{i,i}$, the trace is a *cyclic linear function* which means that it is invariant under the unitary similarity transform such that $\mathrm{tr}(UAU^\dagger) = \mathrm{tr}(U^\dagger UA) = (A)$. The trace $\mathrm{tr}(A|\psi\rangle\langle\psi|)$ is deduced by finding a basis $|i\rangle$ for $|\psi\rangle$ such that the trace can be written $\mathrm{tr}(A|\psi\rangle\langle\psi|) = \sum \langle i|A|\psi\rangle \langle\psi|i\rangle = \langle\psi|A|\psi\rangle$

---

[11] Where $\delta_{i,j}$ is the Kronecker delta function of two parameters $\delta(i,j) = 0$ if $i \neq j$ or 1 otherwise.

[12] This is uniquely defined.

Operator Trace The Hilbert-Schmidt inner product on two operators A,B (also known as the trace inner product) is $(A, B) = \mathrm{tr}(A^\dagger B)$ and is the inner product function on the operator inner product space or operator Hilbert space L_V.

4.2.3 Representing State and Evolution

4.2.3.1 QT Postulates pertaining to State and State Evolution

Associated with any isolated physical system is a complex vector space with inner product, that is, a Hilbert space known as the state space of the system. According to the *first postulate of QT, a system is completely described by its state vector* $|\psi\rangle$, which is a unit vector in this state space. QT does not require one to specify for a physical system what its state space is, nor does it say what the state vector of the system is. The *second postulate of QT states that the state of a closed*[13] *system changes or evolves and this can be described by unitary transformations*. That is, the state $|\psi\rangle$ of the system at time t_1 is related to the state $|\psi\rangle$ at time t_2 by a unitary operator U which depends only on the times t_1 and t_2, thus $|\psi\rangle = U\,|\psi\rangle$. QT also does not specify which U's describe changes in real physical systems, it only suggests that unitary evolution is one way to represent changes.

4.2.3.2 Superposition

A state is said to be in superposition of substates if it can be expressed as $|\psi\rangle = |\psi_1\rangle + |\psi_2\rangle + |\psi_3\rangle$[14]. Given such a system represented by state ψ, it is said that the state experiences 'collapse' when it interacts with its environment (and thereby changes), or any complex external system[15]. In the context of this thesis collapse corresponds to an *ostensive interaction from an external oracle* as elaborated in Chapter 3 in terms of

---

[13] Meaning the system is not affected except as described by the explicit specification of changes in a known QT.

[14] Note that the complex coefficients of each summand is such that the resultant state vector is normalised.

[15] Quantum state collapse is taken here to imply that different elements in the prior quantum superposition are no longer super-posed. The physics or nature of this collapse is not addressed, as in general, the physics of QT do not immediately correspond to IR phenomena, and certainly not applicable in an obvious way. Instead it is said that the higher-ostensive principle corresponds (in QT) to a state collapse, and IN to energy, some other semantical links are explored in Chapter 5.

the principle of higher ostension. The mathematical formalism used to describe the time evolution[16] proposes two different kinds of transformations. Firstly, reversible transformations described by unitary operators on the state space. These transformations are determined by solutions to the Schrödinger equation [Gri02], and since this equation represents the particulars of physical reality the method is said to depend on physical laws of nature. Secondly, non-reversible and unpredictable transformations described by more complex mathematical transformations, examples of these transformations are those that are undergone by a system as a result of measurement and are outlined below.

4.2.3.3 Convex Combinations and the Generation of Probabilities for Events

A state $|\psi\rangle$ is said to be a mixture state or convex sum of states if $|\psi\rangle = \sum w_i |\psi_i\rangle$ such that $\sum w_i = 1$. A mixture state is used to represent a system in terms of a statistical combination of other systems[17]. This type of relationship is not central to this thesis but is key to other applications of QT to IR such as in [MW07b] to model retrieval context which is briefly addressed in Section 4.3.5.4. Whereas convex combinations are not directly employed in this thesis, the idea of generating probabilities for (search) events is present in many applications within this chapter. Where convex combinations are used in this thesis, it is due to their (intrinsic) mathematical convenience rather than their ability to model physical phenomena. This section shows how QT can be used to create models of (observation) events and assign them probabilities, a way which typically employs convex combinations. The QT inspired methods of modelling events with probabilities expressed later in this chapter do not take the typical QT route of using convex combinations and their related theory (and mathematical methods), however, they aid in clarifying the use (in the proceeding) of *density operator* methods. The following outlines the density operator approach of modelling events (search events) that is used for modelling later in the chapter, but is not part of the key modelling approach in this thesis.

During the period of a search session in the traditional perspective, it is the user interests as depicted by a relevance concept, that is evolving. Relevance is then a changing

---

[16] This refers to time evolution of a non-relativistic system [Gri02].

[17] There are further conditions that apply to the use of convex combinations for modelling systems which are are out of scope, a detailed discussion can be found in [BC81]

non-pure state (assuming that more than one document could be relevant) which could be represented as a convex sum of pure states, a convex sum consisting of some pure states and some superposition states, or a superposition of pure states. Consider the first case, with each pure state corresponding to a document represented by the projector onto the 1-dimensional subspace of the document vector. If document vectors $|d_i\rangle$ are normalized on term space with real co-ordinates then their corresponding projectors $|d_i\rangle\langle d_i|$ are idempotent and Hermitian. The state of the system is described by a density operator ρ which is a convex sum of pure-state projectors:

$$\rho = \sum_i w_i \mathbf{P}_i \text{ where } \mathbf{P}_i = |d_i\rangle\langle d_i|,\ wi \geq 0,\ \sum_i w_i = 1 \text{ and } \operatorname{tr}(\rho) = 1$$

Let an observable be described by a Hermitian operator M on $\mathcal{H}$ exhibiting a spectral decomposition consisting of projectors which are pure state projectors or their superpositions (*degenerate eigenvalues*): $M = \sum_i \lambda_i |q_i\rangle\langle q_i| = \sum_i \lambda_i \mathbf{Q}_i,\ \lambda_i \neq \lambda_j,\ \lambda \geq 0$ where each $|q_i\rangle\langle q_i|$ denotes a query[18]. The $\mathbf{Q}_i$ are pairwise orthogonal projectors defining a *refinement*[19] of the identity, $\sum_i \mathbf{Q}_i = \mathbf{I}$. A projector $|q_i\rangle\langle q_i|$ projects onto the i^{th} eigenspace ${}^{Q}\mathbb{E}_i$ of M where $|q_i\rangle$ is an orthogonal basis for that eigenspace. If the eigenvalues are degenerate and the corresponding eigenspace ${}^{Q}\mathbb{E}_i$ has dimension > 2 then any set of pairwise orthogonal vectors spanning ${}^{Q}\mathbb{E}_i$ can be chosen as its basis. As a result, a Hermitian operator has infinite decompositions when its eigenvalues repeat. The set of eigenvalues λ_i are all the possible values this observable can take. The trace function can then be used where the quantity $\operatorname{tr}(\rho\mathbf{Q}_i)$ is the probability that the observable M takes value λ_i in state ρ[20], thus $\operatorname{tr}(\rho\mathbf{Q}_i) = P(Q_i|\rho)$[21], i.e. the probability of the query i given a docu-

---

[18] The 'denoting' here can be symbolic (without explicit matrix representation) for the purpose of designing models. A matrix representation is required for generating numerical results of observations, a simple matrix representation would be to take $|q_i\rangle$ to be the query vector from the vector space model in IR such that $|q_i\rangle\langle q_i|$ is the corresponding *query subspace* in matrix form. M would then denote a sum of weighted orthogonal queries.

[19] See [Gri02] for further discussion on using the method of refining operators including the identity operator to build models of systems.

[20] Note that the a probability is created through a particular setup of geometrical spaces and quantified by the inner product functionals of these spaces. Further, the value $\operatorname{tr}(\rho\mathbf{Q}_i)$ can be written $\operatorname{tr}(\sum_j w_j\ |d_j\rangle\langle d_j||q_i\rangle\langle q_i|) = \operatorname{tr}(\sum_j w_j\ \ \langle d_j|q_i\rangle\ \langle q_i|d_j\rangle)$ employing the cyclic property of the trace function. This can be simplified to $\sum_j w_j\ \langle d_j|q_i\rangle\ \left(\overline{\langle d_j|q_i\rangle}\right)$ as the parameter of the trace is no longer a matrix but a single value. The trace can be further simplified to $\operatorname{tr}(\rho\mathbf{Q}_i) = \sum_j w_j |\langle d_j|q_i\rangle|^2$ since the product of a complex number and its conjugate is a real number.

[21] This way of calculating probabilities (see previous footnote) is said to follow the *Born Rule*, see Ch.

ment collection represented by ρ is calculated by a function that proceeds by multiplying their respective matrix representations. Restricting to simple (non-superposed) pure state query observables, consider a retrieval scenario which may have been given some queries by the user but there is a is a potential query $M' = \mathbf{Q}_i$ yet to be given to the system. The probability value $\mathbf{Pr}(M'|\rho)$ denoting the likeliness of the user posing query M' (i.e. making the observation M') is significant in deciding the probability of relevance of the terms represented by M'.

State can also be represented as a wave function (a unit vector in $\mathcal{H}$), which for the state previously described by ρ would be represented as a normalised superposition of document wave functions $|\psi\rangle = \sum_j |d_j\rangle$, the superposition based representation is the main method employed in this chapter for retrieval modelling and is first addressed in Section 4.3.4.

4.3 Ways of Using QT

4.3.1 Introduction

The QT formalism in particular offers the technique of representing observables in terms of objects in the Hilbert space where operations between these objects can be used to denote physical observations, interactions between subsystems, value measurements and logical relationships [vR04]. This thesis does not address the suitability of using these aspects of $\mathcal{H}$ for designing IR models but instead this section discusses representation issues for documents and relevance values that change over time which it sees as a precursor to deducing how and when an algebra of observables could be effectively used to model search. Further investigations with representing traditional search concepts in vector spaces in the way QT represents and operates on observables can be found in [MW07b, MW07a, Mel07].

---

9. of [Gri02]. A detailed study of doing probabilistic and statistical analysis on the Hilbert space can be found in [BC81].

4.3.2 Using QT vs. Using the Mathematics of QT

4.3.2.1 Use of QT Axioms

The first and second postulates of QT correspond directly to the high-level IN and high-level search from Chapter 3 since the former refers to statics and the latter to the dynamics of a process. Postulates of QT pertaining to measurement and composite systems are not directly addressed by the conceptualisation of search in Chapter 3, however concepts related to these postulates are employed as part of the modelling apparatus in Chapter 5 which also uses the concept of entanglement in Section 5.4.7.1, see [NC00][22]. Other than this, it is the mathematical aspects of QT, particularly the Hilbert space formalism that is employed and in a way that is (in general) detached from the holistic way that QT uses it to model physical phenomena. For example, much of the QT inspired modelling in this chapter is based on using the representation of vectors, subspaces on the complex field, that comes with the $\mathcal{H}$-space theory of QT. Other approaches, upon representing search concepts in QT's mathematical paradigm, exploit a group-theoretic modelling apparatus, which is one of the many related mathematical disciplines that are commonly used within this paradigm (see Section 4.3.7). The use of the mathematics of QT are also further restricted as it is mainly the mathematics pertaining to the first and second postulates that are directly used; for example, product spaces pertaining to composite systems and statistical mixtures (convex combinations) are used only indirectly and in a simplistic way (as indicated in Section 4.2.3.3 for the latter).

4.3.2.2 Using Mathematical formalism of QT in a Specific Way

Much of the mathematical formalism that is employed from QT is also present in other areas. The modelling of state change over time can be taken from Markov models and the modelling of documents and measures between the superposition of documents from basic Hilbert space theory without requiring quantum theory. However, QT provides a *particular way* to use this, so that one can personify/characterize *QT as a researcher agent that uses the abstract device of the Hilbert space and it's own 'quantum-thinking'*

---

[22] As the use of these ideas in Chapter 5 is from a simple conceptual perspective their theory is not further elaborated in this chapter but only briefly alluded to in the next.

to characterise physical processes. In the context of Chapter 3, **the research oracle is the observer using the abstract device of QT to observe a process**, thus, in applying QT to IR the process observed is a search process but using the perspective of QT. Therefore one is learning how to use these abstract devices by *'ostensively' observing the representation* of physical processes and *operational use* of the Hilbert space formalism by QT to derive propositions about these processes. This is similar to saying that the probabilistic models from traditional IR research existed under the name of probability theory prior to IR employing it, but IR employs it in a particular way to model a particular phenomenon and this application is a novel use of probability theory. In this thesis when it is said that QT is being employed what is meant is that, a method of using Hilbert spaces (and related mathematics) is being employed to use Hilbert spaces for something other than what the QT-method uses it for. Clearly if the QT-method of using the Hilbert space and corresponding mathematics were used to represent something close to what QT is representing, that is physical reality, then one can directly import QT-methods. The closer the concept of our abstract search process can be brought to the QT abstraction of a physical process the more of the QT that can be used. This comparison indirectly refers to the conceptual problem in IR, the next chapter addresses this issue more directly whereas the following section addresses different ways of using QT.

4.3.3 Discussion of Conceptual Approaches

4.3.3.1 Introduction

Initially two ways of conceptualising the observation of physical phenomena is presented. These ways are then applied to observing search phenomena followed by a discussion of the properties of search that require to be modelled. In particular, relevance, the user, the system, the evaluation process and measures for evaluating the search process.

4.3.3.2 Discussion of Conceptual Approaches to modelling from QT

Where is the 'quantum character' of IR? The answer is that it is in the assumption that one cannot *know* the user, and therefore cannot fully understand a search process, in the way one cannot completely *know* physical reality. Interpreting IR as a 'quantum physical process' upon it being mapped into the QT domain provides insight into the search process. For the insight to work, the physical interpretation in the QT domain must be conveyable in the IR domain. The IR domain consists of the user, the interface, the interaction language, the system's model of information and the data to be retrieved. In [vR04] there is a reference to [Jau73] who demonstrates an 'interaction protocol' using polarising filters for light. Each filter corresponds to an observable such as the traditional observations of relevance and aboutness. One approach is to address the *effects of a filter* by modelling a *change of phase* which corresponds to an event such as user interaction, system decision making and updating relevance values. For IR, this means modelling effects and not phenomena itself. The phenomena in classical IR terms are user interests in the view of the researcher observing a search process, in the view of the system it is a set of interactions out of which it approximates user interests in terms of relevance concepts, and in the point of view of the user it is system behaviour that includes returning of results.

According to the conceptual model (or semantics) of QT, modelling an effect of a phenomenon or measuring an effect of a phenomenon is equivalent to modelling a measuring device that observes the phenomenon. In IR, modelling of measurement of search effects (which refers to an evaluation) is done by the *interpretation* of user interactions. It is also said that **these two methods themselves interact**. These two models in the language of the concept of higher ostension in Section 3.7 can be conceptually understood as two different abstract devices for interpreting a scenario. Further, from the perspective of an external researcher (studying these methods) they denote two researchers. These can be conceptually related by letting R_w denote the wave or phenomena modelling method and R_m denote the method of making observations and measurements[23]. A search activity corresponds to a measurement, the measuring equipment or IR evaluatory measures act as a filter or observer for the search phenomena that corresponds to the 'characteristic wave'[24] in QT. The search activity itself can be said to be an observable entity with relevance, and aboutness as observables (as in [vR04]) that together form the *search observable*, hence there is the notion of a compound (or summed) observable. User interaction (exter-

---

[23] R_m and R_w are abstract devices and also researchers in the conceptualisation of Chapter 3.

[24] In accordance with physical semantics this is more commonly termed as the characteristic electromagnetic wave.

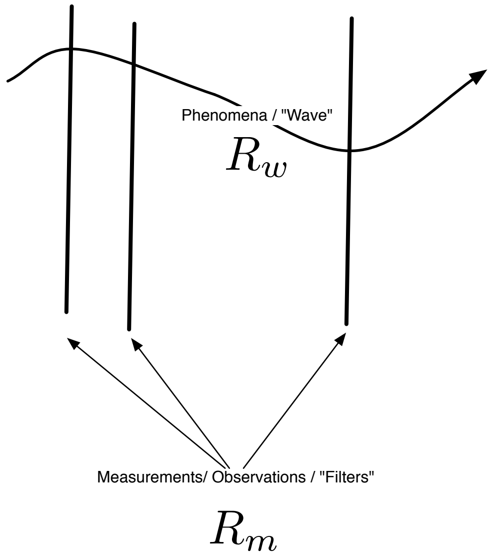


Figure 4.1. The Wave method R_w and Observation/Measurement method R_m for modelling phenomena

nal to system) and system decisions (internal events) correspond to observing something about the 'phenomena' or search activity (or search sub-activities[25]). Measuring aboutness/relevance correspond to observables as do the performance or observation of an user interaction or any other activity. Applying these observables corresponds to measurement of the wave using filters, Fig. 4.1 illustrates this.

A metaphor of these two researchers or modelling techniques R_w and R_m is that of the interaction between two people as viewed by an external observer[26]. R_m is *travelling* for a period of the time that denotes the period of a search process or of a process in which the phenomena to be measured exists. R_m holds properties that are said to exist prior to measurement since it corresponds to a complete model[27] of the phenomena[28]. R_m then interacts (or is influenced) by R_w causing it to form a direct 'opinion' (observation) about R_w and an indirect opinion about the phenomena corresponding to the model. R_m does not know the specific details of R_m, thus the person R_m knows that it 'experienced something'[29] and can describe this in its own terms but not in terms of the model or in the conceptualisation of person R_w. Following this interaction, R_w is also affected and

---

[25] Such as clicking a button on an interface.

[26] Which in this case is the IR researcher or reader of research.

[27] Thereby possessing complete knowledge of it.

[28] This pre-existing design or knowledge of measurements is a QT semantic.

[29] In accordance with the concept of experiential gap introduced in Section 3.7 let this event be known as an *information experience*, this is elaborated through an application in Section 6.5.

makes opinions about the interaction, this corresponds to phase change for a QT model (of R_w). R_w also has no knowledge about R_m in an objective language[30] but only in terms of its concepts, such as phase change. Apart from this 'meeting' (interaction) each person/researcher/model has no knowledge of one another.

QT is an Abstract Device of a Researcher in High-Level Search The model given in Section 4.3.4 is directly modelling the phenomena in a system that itself measures the phenomena of user interests. Hence it is a direct model of system's relevance phenomena so **$\mathbf{R_{w-thesis}}$ models the phenomenon $\mathbf{R_{m-system}}$ which it perceives as $\mathbf{R_{w-system}}$, and $\mathbf{R_{m-system}}$ models $\mathbf{R_{w-user}}$.** In this way, both approaches are related on a conceptual level[31] as exemplified by Section 4.3.4. According to the general researcher (who sees the overall interaction), IR research corresponds to both these two approaches. The two approaches in turn correspond to two different types researchers and observers.

The observation scenario where R_m and R_w relate can be represented by Fig. 4.2 where R_w is a 1-dimensional line representing phenomena that is in a 2-dimensional world according to R_m, R_m is represented by a 2-dimensional plane in a 2-dimensional world (according to it). The observer $R_{observer}$ is in a 3-dimensional world represented by the cube in which R_m and R_w are subspaces $R_{observer}$ is able to clearly tell how the lesser dimensioned entities interact. Another entity $R_{external}$ is said to occupy an N-dimensional world ($N > 3$) in which $R_{observer}, R_m, R_w$ are subspaces.

R_w and R_m move past one another in $R_{observer}$'s world such that they both occupy[32] and affect each other, but R_w has no idea of R_m as it is out of its scope, and similarly for R_m and $R_{observer}$. To the external observer $R_{external}$, $R_{observer}$ is an abstract device or way of looking at R_m and R_w. In classical IR there is the user, system, search session, researcher (who designs/creates/theorizes). In our formulation of IR, according to the principle of higher ostension, a search session characterizes the user and system as one compound observable, and there are two methods: phenomena modelling and measurement, by which

---

[30] In a form that can be understood by a third party that did not observe their interaction.

[31] And therefore it appears to avoid the debate resulting from the statement "No elementary Phenomenon is a Phenomenon until its observed" [Whe80] (as mentioned in [vR04]) as it can be said that we do not have an elementary phenomenon and instead QT is 'custom-used' in a way that ignores much of the semantics pertaining to physics and physical reality).

[32] Like '2D Cartesian planes' "flying around intersecting with each other".

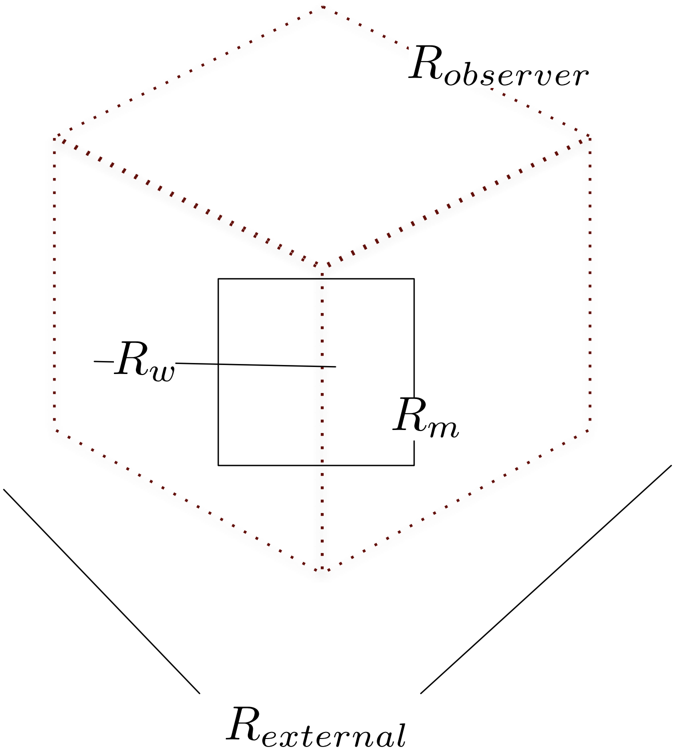


Figure 4.2. Geometric Metaphor for relationship between phenomena, measurement and observation approaches

to model search. Generalising in terms of Fig. 4.2, $R_{observer}$ *corresponds to QT* as it is the abstract device being employed. In this way *the search process is the phenomena or wave that is travelling (changing state)*. The conceptualisation of search in Chapter 3 states that the search process is not only a process but a process by which all other processes are known. Interpreting this concept using the physical semantics of QT, it can be said that the search process is a phenomenon that is itself a measurement of other phenomena. Following from this, in terms of phenomena, one can say that the search process moves around (and is therefore like R_w) where *its 'position' corresponds to a set of search observables* in the space R_{m_2} which is a 'parallel' of R_m[33].

Corresponding to R_w and R_m there are two representation paradigms, the **wave function method** and the **measurement method**, the former implies modelling phenomena directly thereby corresponding to R_w. It is similar to the method of depicting the anomalous

---

[33] Note the peculiar concept here, a search process, search phenomenon or search 'wave' (all conceptually equivalent and corresponding to R_w) has a 'position' which does not indicate its locality in space but indicates *the nature of what it measures/observes of other phenomena* (such as user interaction in the case of low-level/traditional search).

state of knowledge from [BOB82b] and corresponds to direct modelling. The measurement method pertaining to density operators and measuring phenomena or effects thereby corresponding to R_m, it also corresponds to the traditional ostensive retrieval approach of 'capturing information need', and was explained in Section 4.2.3.3, this method is not employed for modelling (except indirectly) in this thesis.

4.3.4 A Technique for Modelling Simple Searches

4.3.4.1 Introduction

This section presents formal constructs borrowed from QT to represent a cognitive model of an ostensive agent (see Section 2.3 and Section 2.4). This is approached by treating the paths in the ostensive model (OM) as a direct representation of 'its cognition' (see Section 2.4.5). The QT based representation in this chapter of ostensive search paths suggests novel ways of analysis of the user interactions (denoted by the paths). These ways of analysis correspond to the *single concept properties* of 'association strength', 'uncertainty of association' and 'association reasoning' (Section 2.4.2), that is, ostensive search paths can be analysed by comparison with one another in the context of these concepts. In particular, the discussion in this section results in the *utility measure* which can be used to measure the 'strength of association' between search paths.

The utility measure is presented in Section 4.3.4.6 and is further discussed towards the end of the chapter in Section 4.3.4.8 in terms of its relation to correlation measures. The discussions of these measures aim to provide a comparative interpretation of *utility* and show that it is an informative measure and corresponds to a useful semantic for retrieval. Discussion in Section 6.4 shows how this measure can be used to generate new retrieval models. This thesis does not create a model for complex cognitive changes occurring in the user during search but instead it approaches the simpler problem of modelling the simple 'cognitive agent': the system. A framework is developed with the Consistent Histories interpretation[34] of QT [Gri02].

---

[34] The suitability of a particular QT interpretation over another for modelling IR is not addressed, and the thesis does not actually address quantum histories except briefly in Appendix C, instead it presumes the interpretation of QT in [Gri02].

4.3.4.2 Relevance States

This section presents a model for search in the QT framework by modelling the change of relevance from the perspective of traditional search. Firstly, following from Section 2.4.3 two states are defined, the **system** is represented by a description of the users information need which is *inscribed* in a state $|\Omega\rangle$, and there is the equivalent system state, the *relevance state* $|\psi\rangle$. $|\Omega\rangle$ represents the user's cognition and $|\psi\rangle$ the 'cognition' of the ostensive agent (or system). $|\Omega\rangle$ varies over the retrieval process and $|\psi\rangle$ varies according to observations made about $|\Omega\rangle$. The information need state $|\Omega\rangle$ is a very general property of the user's cognitive state and cannot be represented with certainty.

The aim of a retrieval system is to fulfil the users information need, requiring minimal effort on the part of the user [Zip49]. Effort in the ostensive retrieval sense is proportional to the weighted number of interactions between the user and the retrieval system. Without a representation for the users IN state $|\Omega\rangle$ one is restricted to the system's view of this state as inscribed in the ostensive agent's cognitive state $|\psi\rangle$ which will also be referred to as the *relevance state*[35]. This means that the framework is *recording* the search process by modelling how it changes the system, or *relevance state* $|\psi\rangle$.

At each stage the retrieval system progresses by updating $|\psi\rangle$ and visually by changing its interface to reflect its new state. Additionally it is required that this visual change be in anticipation of further interactions so as to *reduce effort*. This is a high level description of the correspondence between $|\psi\rangle$ and $|\Omega\rangle$ in terms of user interaction, relevance, IN and change in IN[36]. This is a *low level retrieval strategy*, an abstraction of what already happens in the ostensive model (OM).

4.3.4.3 Representation of Search paths

Formally the path model in OM representing an ostensive search process can be formulated as a single source directed graph $\overrightarrow{OG}$ where each vertex denotes an object with which the user can interact. Let $\overrightarrow{OG} = (V, E)$, there is an edge $(v_i, v_j) \in E$ if and only if user

---

[35] In terms of researchers (Section 4.3.3) this scenario is from the perspective of $R_{w-thesis}$ (doing a high-level search) observing $R_{m-system}$ which appears as phenomena as denoted by $R_{w-system}$.

[36] This is in line with the discussion of the *purpose* of a retrieval agent (Section 2.4).

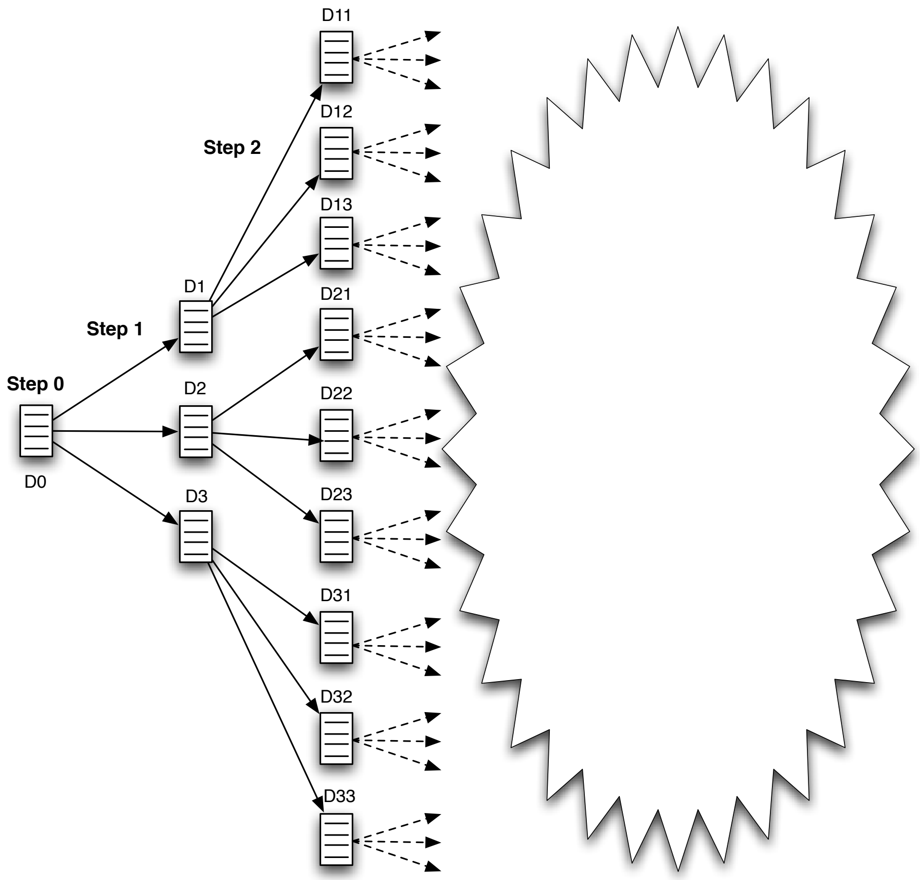


Figure 4.3. The ostensive browsing process represented by a graph

interaction with v_i *caused* the system to *recommend* v_j, these correspond to the candidate next steps illustrated in Fig. 2.4. The ostensive graph OG defined above can be visualised from Fig. 4.3 which has a constant number of recommendations at each stage that is denoted by the degree of the graph δ (the degree is three in the diagram).

The framework models search according to the OM quite closely at this stage, assuming an initial document and corresponding candidate next steps. Let $|_t\psi\rangle, t \geq 0$ denote the relevance state of the system at time/step t, then step 1 is a passive step corresponding the assumption of an initial document, and assume that a set of relevance judgements from the retrieval system in terms of probabilities as is the case in [Cam00, CvR96, URJ03] where the binary voting model is used. In further discussions a discrete time scale is assumed which increments with user interactions or steps in the retrieval process allowing the time to be interpreted in terms of steps. Proceeding with the document representation, the representation of relevance states are defined, initially giving two examples as expressed in

Equation 4.2 and Equation 4.3 fixing the degree of the graph or number of candidate next steps to δ (without loss of generality). Such a document representation is exactly that of the vector space model as although a *finite dimensional* Hilbert space $\mathcal{H}$ is being used only real numbers are being employed to represent term weights thus the mapping back to the traditional vector space is trivial. In the quantum formalism each document vector whose lengths are normalised to unity are known are *state vectors*. They are represented using the elementary bases corresponding to terms in T dimensions, see Equation 4.1. At this stage real weights such as those generated by the tf-idf methods are assumed, they will from now on be referred to *static weights*. The only difference at this stage to a vector space retrieval model is the use a Hilbert space that allows for complex values which are yet to employed. This difference is an important one as QT systems are modelled in such a space but at this stage the use of $\mathcal{H}$ is intentionally limited. A discussion of a representation technique different from the one used in this section that uses complex numbers can be found in Appendix B.

$$|d_i\rangle = \sum_j w_{i,j}\,|w_j\rangle, \text{ where } \quad w_{i,j} \in \mathbb{R} \tag{4.1}$$

$$\begin{gathered} |{}_1\psi\rangle = \sum_{i=1}^{\delta} {}_1p_i\,|d_i\rangle \\ {}_tp_i \in \mathbb{C} \end{gathered} \tag{4.2}$$

$$|{}_2\psi\rangle = \sum_{i=1}^{\delta} {}_2p\,{}_{1i}\,|d_{1i}\rangle + \sum_{i=1}^{\delta} {}_2p\,{}_{2i}\,|d_{2i}\rangle + \sum_{i=1}^{\delta} {}_2p\,{}_{3i}\,|d_{3i}\rangle + |{}_1\psi'\rangle \tag{4.3}$$

In the above equations all *possible* user interactions are represented with objects by a superposition of object/document states (the ($|d_i\rangle$'s)) with which the user can interact. Such a representation was inspired by the idea of a state of superposition for cognitive concepts in [AG02] and discussed in the previous chapter. The probabilities of relevance for each document with consideration for anticipated relevance in the next step are encoded in the ${}_tp\,{}_i$'s so that the probability of the user interacting with $|d_i\rangle' s$ at time t given $t-1$

interaction activities can be extracted from the ${}_tp_i$'s. At step one (see Fig. 4.3) the user has the choice of interacting with only δ documents (three in the figure) each with the probability inscribed in ${}_1p_i$, and in step two in where the user has interacted with one of these, the result is a further branching the states increase in Equation 4.2.

Each summand in Equation 4.2 is to be interpreted as a possible path that can be taken in the ostensive graph if the respective item from $|{}_1\psi\rangle$ was interacted with in the preceding step. $|{}_1\psi'\rangle$ denotes the objects on screen as observed during time/step 1 that are still considered potential interaction points in step 2, thus this is just $|{}_1\psi\rangle$ with the probabilities updated to reflect the change in IN and the corresponding change in the relevance state is denoted by $|{}_1\psi\rangle \rightarrow |{}_2\psi\rangle$. According to Fig. 4.3 if the user clicked on d_1, a new set of recommendations would be generated so that in step 2 one can interact with any of d_{11}, d_{12}, d_{13} represented as states $|d_{11}\rangle$, $|d_{12}\rangle$, $|d_{13}\rangle$ or any of $|d_2\rangle$, $|d_3\rangle$ which are accounted for in $|\psi_2\rangle$ by the $|{}_1\psi'\rangle$ term. In summary, the representation of the relevance state is a weighted superposition of all documents with which the user can interact. Let $|{}_t\psi\rangle = \sum_{i\in\, {}_t\chi \subset V} {}_tp_i\, |d_i\rangle$ where ${}_t\chi$ is the set of all items with which the user can interact at time t, these are the set of possible future states as in Section 3.4[37]. In the traditional case ${}_t\chi \subset V$ where V is a subspace of the document space. The relevance state can also be expressed with a complex numbered phase in which the probability of relevance of a document is embedded $|{}_t\psi\rangle = \sum_{i\in\, {}_t\chi \subset V} {}_te^{\jmath\theta_i}\, |d_i\rangle$ where $\theta_i = f(p_i)$, i.e. θ_i is a function (or encoding) of the probability of item i. As a document can be expressed in terms of words and $|{}_t\psi\rangle$ is a sum of documents that can be represented in terms of words $|{}_t\psi\rangle = \sum_{i,j} (w_{i,j})\, {}_te^{\jmath\theta_i}\, |w_j\rangle$[38] or without loss of generality if the relevance state is represented in matrix form then ${}_t\psi = \sum_{i,j} (w_{i,j})\, {}_te^{\jmath\theta_i} |w_j\rangle\langle w_j|$[39]. Let $\mathbf{Pr}(Rel|x)$ denote the probability of relevance of the object x which can be a document, a term or some other construct. The form $\mathbf{Pr}_x(Rel)$ denotes either the probability of relevance of an item x or an item with index x (in a set of items) depending on the context . The next section explains how the relevance state is manipulated to indicate the change in the probabilities

---

[37] Assuming the modeller knows the future possibilities.

[38] This representation assumes do not change over time but that only the user's interest in them changes; otherwise, each term could be written with a time parameter ${}_tw_{i,j}$, changing documents are not considered in this thesis.

[39] Symbols of weights have been kept the same but the weights are not necessarily the same in both representations.

of relevance of documents.

4.3.4.4 Evolution of relevance states

In quantum physics the evolution of a system from one state to the next, or the dynamics of the system, can be represented by an unitary operator U in $\mathcal{H}$. Unitary operators are required for representing dynamics that are reversible. In general, dynamics of a system are not necessarily reversible meaning that once a physical system changes from one state into another it is presumptive to think that it can change back to a previous state and if so that one know how this happens in terms of operators on $\mathcal{H}$. In general, for ${}_{t+1}\psi = T_t\psi$, T may not be known or if it is known and is deducible by matrix operations it may not correspond to a physical semantic thus may not correspond to reality. Also T^{-1} may not exist which on the mathematical level means that there are no simple operations to get back to the original state[40]. In general if a state changes by operations T_1T_2, it is not necessarily the case that the result state is equivalent to changes by operation T_2T_1, thus operator application is not generally commutative $T_1T_2\left({}_t\psi\right) \neq T_2T_1\left({}_t\psi\right)$. In IR, a real user's *search experience is not commutative* as interaction with documents in a particular order could lead them to interact/behave differently in the future than if the initial interaction was with documents in a different order. Commutativity in search is at most a special case of symmetry in the effects of sub-processes[41] that are part of the greater search process. The non-commutativity of real search processes was also one of the reasons for proposing abstract (and simulated) users `AU` as being necessary (as in Section 3.7) for addressing the definition problem and conceptual problems. This is since real (non-commutative) users would be difficult to analyse whereas (simulated) commutative exhibit more regularity, are (in general) easier to model and can approximate real users. One of the results of such a conceptual proposal is exemplified in Section 4.3.7 where symmetry induced by commutativity conditions are used to suggest particular simulation designs.

In order to capture changing information need in the relevance state the unitary operator

---

[40] If in reality (semantically) there exists an operator that reverses a state transition then it not necessarily true that this can immediately be represented in the Hilbert space in a consistent way. For example, if a user can 'un-think' their interests then it is not the case that this 'un-thinking' can be represented by a the model describing user interests.

[41] A simple case is when a checkbox is clicked (checked) and then clicked again (unchecked).

is applied to a state vector (such as $|{}_t\psi\rangle$) at time t resulting in the succeeding state vector. The $|{}_t\psi\rangle$ vectors evolve from one time to another, going from $|{}_t\psi\rangle$ to $|{}_{t+1}\psi\rangle$ by the mathematical means of $|{}_t\psi\rangle$ getting multiplied by a matrix ${}_{t+1}U(t \geq 1)$[42]. In order to represent the logical evolution of our relevance state the unitary matrix of Equation 4.4, a diagonal matrix[43] with dimensions $N \times N$ is required where N denotes the number of unique terms in the collection.

$$
{}_{t+1}U = \begin{pmatrix}
{}_{t+1}\omega_1 & 0 & \multicolumn{4}{c}{\dotfill} \\
0 & {}_{t+1}\omega_2 & 0 & \multicolumn{3}{c}{\dotfill} \\
0 & \ldots & {}_{t+1}\omega_3 & 0 & 0 & \ldots\ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \ldots & 0 & \ldots & 0 & {}_{t+1}\omega_N
\end{pmatrix} \tag{4.4}
$$

Unitarity Conditions:

$$U^\dagger = U^{-1}, U^\dagger U = UU^\dagger = U^{-1}U = UU^{-1} = I$$

$${}_{t+1}\omega_i = ({}_{t+1}a_i) \pm \jmath\sqrt{1 - {}_{t+1}a_i^2}\,\jmath = \sqrt{-1}$$

The matrix ${}_{t+1}U$ is represented on a term basis such that ${}^{w}_{t+1}U = \sum_j ({}_{t+1}\omega_j)\,|w_j\rangle\langle w_j|$ where ω_j are the eigenvalues of U that lay on the complex unit circle, ${}_{t+1}\omega_j = {}_{t+1}e^{\jmath\theta_j} \in \mathbb{C}$. Inner products and therefore lengths of vectors are invariant under unitary evolution. This means that the geometrical effect of applying U to an operator (state, or matrix) is a projection onto the eigenspaces of the operator followed by a rotation within the respective eigenspace. In IR terms this means that one enters into the space of a particular term then changes its value, which due to the value being a complex number, has a geometric effect is visualisable as a rotation in $\mathbb{C}$ by θ.

States can also be modelled such that the basis are not terms but documents represented as an orthogonal basis. The assumption here would be that each document is an unique object, and thus ${}^{d}_{t+1}U = \sum_i {}_{t+1}\omega_i |d_i\rangle\langle d_i|$ where each ω_j denotes the change in relevance of document j in time $t+1$. In order for ${}_{t+1}U$ to work as an operator that shifts the probabilities of the objects of its basis (be it terms, documents or other items) it has

---

[42] The time index of U starts at one as U represents a state change which implies that a state exists at $t = 0$ for U to change and therefore the first change has to be at $t > 0$.

[43] A square matrix in which all values outside the diagonal are zero.

to initially (at $t = 1$) be set such that the elements of ${}_1U$ are ${}_1\omega_i = {}_1e^{\jmath\theta_i}$ where $\theta_i = f\left(\mathbf{Pr}(Rel|d_i \text{ or } w_i \text{ or } \ldots)\right)$[44], i.e. the elements must hold the probability of relevance of the items in the basis or a 'encoding'/function thereof[45]. This is so that the state being changed by U can be assigned its initial dynamic weight on first application of an evolution operator.

Change of Basis Each $|d_i\rangle\langle d_i|$ can also be represented in terms of $|w_j\rangle\langle w_j|$ such that the unitary change operator can be written[46]:

$$ {}^{d\to w}_{t+1}U = \sum_i {}_{t+1}\omega_i \left(\sum_j w_{i,j}|w_j\rangle\langle w_j|\right) = \sum_{i,j} w_{i,j}\left({}_{t+1}\omega_i\right)|w_j\rangle\langle w_j| \qquad (4.5) $$

The operator ${}^{w}_{t+1}U$ can similarly be represented in a document basis by setting:

$$ {}^{w\to d}_{t+1}U = \sum_i \left(\sum_j w_{i,j}\left({}_{t+1}\omega_j\right)\right)|d_i\rangle\langle d_i|, \text{ assuming normalisation} \qquad (4.6) $$

where each $u_{i,i}$ (the i^{th} diagonal of U) denotes *the average extent (over all terms) to which document i is relevant to the current information need that is specified by probabilities of relevance of terms embedded in ω_j*. The complex numbered entries ${}_{t+1}\omega_i$ of U together satisfy unitary conditions and each ${}_{t+1}a_i$ in Equation 4.4 is a real number indicative of the change of the probability of relevance $\mathbf{Pr}\left(Rel|_t d_i\right)$ of $|d_i\rangle$ based on observation of user interaction at step t. The ${}_{t+1}\omega_i$'s are bounded by the half circle of radius one in the complex plane, and their argument is $-\frac{\pi}{2} \leq {}_{t+1}\theta_i = \arg({}_{t+1}\omega_i) \leq \frac{\pi}{2}$ radians[47]. The rotations they induce on the coefficients in $|_t\psi\rangle$ can be interpreted as increase or decrease of relevance of the corresponding document with the overall change in all documents summing to zero. Representing changes in this way means that the system interprets relevance in the documents shown between steps as just 'changing hands' over visible[48] documents as

---

[44] The unit of information can be a term, document, sentence, paragraph or otherwise.

[45] An example encoding of probabilities of relevance can be found in Section A.1. One use for employing a function instead of actual probabilities is to force the 'shifting of probabilities' on multiple applications of the unitary operator to resemble movement in the complex plan that has some visual (geometrical) significance, see Section 4.3.4.5.

[46] It is assumed here that all the necessary normalisations are done such that $\sqrt{\sum_{i,j} w_{i,j}\left(\omega_i\overline{\omega_i}\right)^2} = 1$.

[47] For a discussion of several ways to use unitary operators (albeit not in the QT framework) to represent documents see [Hoe03].

[48] Refers to visibility on the ostensive graph.

opposed to 'disappearing'. This is equivalent to saying that the probabilities of relevance of all the items forming the basis sum to one.

Upon arbitrary step evolution of the relevance state, given the diagonal form of the unitary matrix one gets $|_t\psi\rangle = \sum_{i\in\ _t\chi\,\subset\, V} e^{\jmath\,(_t\Theta_i)}\,|d_i\rangle$ as derived in Appendix A which is the relevance state in terms of the *relevance judgements* $_{t+1}\omega_i$ at successive times; $_t\Theta_i$ encodes the resultant dynamic weight at t. A further representation for the relevance judgements is introduced next, which helps to illustrate evolution of the relevance state in a geometrically intuitive way.

Unitary Evolution does not Predict Relevance state evolution[49] has the geometric effect of a successive number of rotations in the complex plane for each item $|d_i\rangle$ with the rate of change of magnitude of each rotation $(_{t+1}\theta - {}_t\theta)$ reflecting the rate of change of relevance of that item. An important point here is that the dynamics of the system change *per interaction* so one cannot necessarily deduce $_{w>t}U$ and $_{z>t}U$ from $_tU$ by performing operations (corresponding to some interactions) on previous evolution matrices since the prior dynamics (generally) no longer apply[50]. This re-iterates the particulars of the definition of state of a quantum system by the second postulate of QT[51], as to find an equation for evolution one would need a model for relevance evolution, which corresponds to modelling dynamics and to theories about the nature of physical reality in the QT framework; Appendix C provides some commentary about using QT's physical semantics to model relevance. However, Section 4.3.7 presents some initial (simpler) alternatives to modelling the change pattern of relevance over time[52].

The $|d_i\rangle$ are represented as real valued vectors in T dimensional term space but when each $|d_i\rangle$ in $|_t\psi\rangle$ is expanded to represent it in term space, $|_t\psi\rangle$ becomes a complex valued vector in the space. Hence the relevance state transitions can be seen geometrically on the term

---

[49] The encoding of which is detailed in Appendix A.

[50] In physics the continuity of properties or the representation of change of a property by equations is possible, however, in this case there is (in general) no such continuous pattern since one is dealing with cognition.

[51] That unitary matrices can be used to represent change but QT does not give us a way to predict or deduce this change, for which additional theories are required.

[52] For conceptual completeness, note that this also corresponds to the changing dynamics of the system view of users IN as discussed in Section 2.3.

space as scaled rotations at each stage, this is a reversible operation[53]. In this section, it was shown that information need can be symbolised and recorded in a QT framework, in a way such that it can be said that the the framework 'takes notes' about the search.

Static and Dynamic Components of a State Interpreting a change characterised in terms of words in the form of documents (in a document basis) and applying ${}^{w\to d}_{t+1}U$ to a relevance state expressed in terms of documents which corresponds to changing the basis of the representation of relevance states from a word basis to a document basis, would result in updating both the states' static part corresponding to the length of complex coefficient (in polar co-ordinates) and dynamic part, corresponding to the argument of the complex coefficient (angle of complex coefficient in polar co-ordinates). The effect of this is that the prior relevance value of the document, the θ_i in $|{}_t\psi\rangle = \sum_i (w_i)\,{}_te^{\jmath\theta_i}\,|d_i\rangle$[54] is modified to represent the average (over all terms in that document) change of relevance of words in each d_i. The new length part w_i of each document matrix coefficient (deriving from $\sum_j w_{i,j}$) can be interpreted as the average extent[55] to which collection terms are represented in the document.

${}^{d\to w}_{t+1}U$ is applied to a relevance state expressed in word basis then the angle part of the coefficient of each word matrix $|w_j\rangle\langle w_j|$ is the average change of probability of relevance of the word at a particular time over all documents and this is zero. The static weight or length denotes the average representation of that word in all documents, this is implied due to probabilities needing to add to 1[56]. In this chapter, probabilities of relevance as taken are assumed to represent IN, this is challenged and generalised in the next chapter where various alternative quantities are presented.

---

[53] This means that the operation of rotation, say by θ, has an inverse that rotates by $-\theta$.

[54] The representation of relevance state in terms of documents only where $w_i = 1$.

[55] Or more formally, a function of the average of tf-idf values in a document.

[56] This can also be used to support the semantic (partially expressed in Chapter 3) that IN should be perceived qualitatively and quantitatively as that which does not 'increase or decrease' over the set of all components of a search, however, it can increase/decrease relative to a sub-component, as addressed in the conceptualisation of Chapter 5. Therefore, it is similar to the physical concept of energy. It is said that the components or sub-components (as formalised in the next chapter) of a search correspond to the physical systems and sub-systems where the overall change in energy of a closed system is zero whereas for its sub-systems it is not.

4.3.4.5 Visualising in Polar Co-ordinates

Given a representation $\sum_{i,j} w_{i,j} e^{\jmath\theta_j} |w_j\rangle\langle w_j|$ let $\nu(\psi_a, \psi_b)$ denote the inner product between states:

$$ {}^t\nu(\psi_a, \psi_b) = \langle \psi_a | \psi_b \rangle = \sum_{i,j} {}^a w_{i,j} {}^b w'_{i,j} e^{\jmath({}^a\theta_j - {}^b\theta_j)} \text{ see Fig. 4.4.} \tag{4.7}$$

The angle between the subspaces in the above inner product is:

$$ f(\arg({}^w\nu(\psi_a, \psi_b))) = \sum_j \mathbf{Pr}_a(w_j) - \mathbf{Pr}_b(w_j) \tag{4.8}$$

which denotes average change of relevance over all terms and this should be zero as:

$$ \sum_j \mathbf{Pr}_a(w_j) = \sum_j \mathbf{Pr}_b(w_j) = 1 \tag{4.9}$$

If change was not represented at the term level then the argument of this inner product would denote the average difference in relevance of a term[57]. Fig. 4.4 visually represents[58] the process of summation of complex coefficients of relevance states (or components of relevance states) where each arrow represents the effect of adding a summand to a cumulative sum in the calculation of $\langle \psi_1 | \psi_2 \rangle$[59]. The shape of the path from $(0, 0)$ to $(1, \sigma)$ could present a useful visual summary of the inner product operation between states if there was a meaningful order on the set of components of states; thus, the path can show what set of changes are required for one state to transform into another.

Ordering of Bases Thus if $|d_i\rangle\langle d_i|$ or $|w_j\rangle\langle w_j|$ corresponded to a semantic order $([d_i], \prec)$ or $([w_j], \prec)$[60] respectively then the visualising of summands of calculations between states in bases $|d_i\rangle$ or $|w_j\rangle$ respectively could be used. For example, a document set can be ordered according to a ranking corresponding to prior investigations of relevance to some set of

---

[57] When the relevance state is represented in terms of documents and then re-represented in terms of words then the angle part of the coefficient of word matrices are sums of the probabilities of relevance of all the documents the word was in, and since a word may not be in a particular document the sum of this sum of angles that would be present in an inner product between states would not necessarily be zero.

[58] This is simply a sum of complex number on an Argand diagram.

[59] Note that ψ_i refers to an arbitrary state labelled i, ${}_t\psi$ refers to a state at time t and the pair ${}_a\psi_1, {}_b\psi_2$ refer to two different states at different times.

[60] Recall that $(A, \prec)$ denotes a binary order relation on the set A (see [Ros99]).

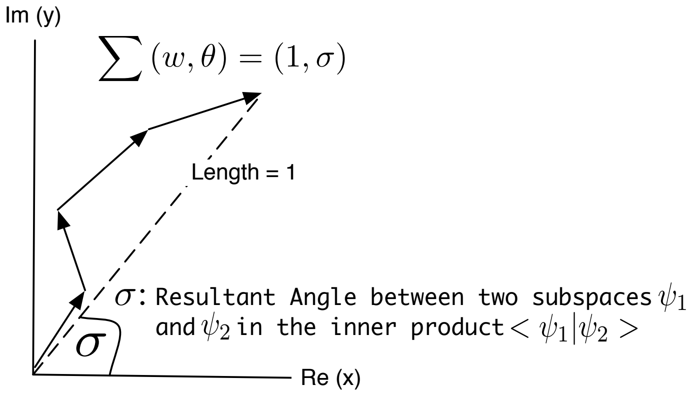


Figure 4.4. A graphical representation of the result of adding a set of complex numbers in polar-co-ordinate form that represents the result of operations of the relevance state

queries, in which case the 'change-path' in Fig. 4.4 can be used to immediately generate statements such as *"..as we get more into a search process the relevance of documents from d_{30} to d_{37} increase"*, indicating that a particular IN that was not very prevalent in prior searches seem to be rising in rank in this one. For a term basis, an order can denote a linguistic semantic such as the word-frequency over a particular collection thus allowing propositions of the form *"..user interests are increasing for words of a particular popularity level [popularity according to some other collection] as search goes on, so it looks like this user would/would not benefit from using this other collection based on his search behaviour on this collection.."*. This visualisation technique and visual symmetry on the Argand diagram is useful for representing symmetrical relationships between term weights, especially between the tf and idf values of terms, this is further elaborated in Appendix B.

The above visualisation technique is perhaps even more useful in interpreting state evolution which corresponds to the polar-rotation of each coefficient and can be equivalently represented in the form of Fig. 4.4, and if an order on the bases existed, statements about patterns of IN change can be suggested[61]. Using the same notation one could analyse the change of a weight over a search session that involved multiple interactions characterised by changes of relevance embedded in U's, and this can be represented as a moving object in the complex number space (as depicted by the arrows in Fig. 4.4), this would be useful for designing simulated changes for searcher simulation which is covered in Section 4.3.7.

---

[61] As the visualisation in polar co-ordinates can correspond to changes in IN if the basis are ordered one can compare it to semantical diagrams of IN change in [Cam00].

4.3.4.6 Comparing search paths

This section shows that the inner product between two states defined on $\mathcal{H}$ can be interpreted as state similarity in the way the dot product between document vectors refers to document similarity. In order to reach this interpretation one requires to modify the intuitive encoding of the probability of relevance in the above section to one that gives a more reasonable result to our calculations (see Appendix A). At this stage it should be elucidated that the two states involved in the inner product can represent search paths during the same session, a state with an anticipated search path that is being evaluated, or two general search states[62]. Define ν_0 as the **Utility function** on a pair of states/searches, it is defined as follows:

$$\begin{aligned} \text{Let } \nu(d_i, d_j) &= \left(\mathbf{Pr}\left({}_td_i\right) - \mathbf{Pr}\left({}_td_j\right)\right)\langle {}_td_i|{}_td_j\rangle \\ \text{Let } \nu_0 = f\left(\left\langle {}_t\psi \middle| \widehat{{}_t\psi}\right\rangle\right) &= \frac{\sum\limits_{i\in{}_t\chi, j\in{}_t\chi'} \nu(d_i, d_j)}{\sum\limits_{i\in{}_t\chi, j\in{}_t\chi'} \langle {}_td_i|{}_td_j\rangle} \\ -1 \leq \nu_{ij} &\leq 1 \end{aligned}$$

$$\nu_0 : \begin{cases} = \infty & \sum\limits_{i\in{}_t\chi, j\in{}_t\chi'} \langle {}_td_i|{}_td_j\rangle = 0 \text{ (This is a special case see the proceeding).} \\ \in [-1, 1] & \text{otherwise} \end{cases} \tag{4.10}$$

The information need of an user associated with one state need not be the same as the other state; since ${}_t\psi$ is normalized the real part of the similarity value can be interpreted as the average distance between documents in each state, while the imaginary part is a weighted average. Hence two states are distinguished according to a combination of the similarity between their documents and the relevances of these documents. The utility function captures this distinction in Equation 4.10 and presents itself as a measure for quantitatively comparing states based on dynamic and static search information[63].

---

[62] Even more generally it could correspond to knowledge states, since a search state corresponds to the knowledge of an abstract device or researcher according to Chapter 3

[63] The prior probabilities in classical probabilistic models account for the static information represented here. The utility function is otherwise quite different from the classical models, firstly in the way it

4.3.4.7 Interpretation of ν_0

In the prior section a general measure by which to compare two search paths was given, and is derived from the argument of the inner product as shown in Equation 4.10. In order to interpret ν_0 one must first understand its constitutes. Each summand $\nu_{i,j}$ is the cosine similarity (as document lengths are normalized to one) between documents *i* and *j* multiplied by the differences in their probabilities of relevance. Although the dot product is a stronger measure of similarity on a semantic space like in latent semantic indexing [DCC$^+$03], one can interpret its value as an indication of *how well the documents represent each others concepts (terms)*. The subtraction of probabilities indicates how different the documents are according to their perceived importances/relevances.

The value $\nu_{i,j}$ tells us in which document the concepts common to both documents are more needed[64] and by how much. If in $\nu_{i,j}$, ${}_td_i$ and ${}_td_j$ are about the same concepts while ${}_td_j$ is more relevant then the value of $|\nu_{i,j}|$, according to **only** these two documents, gives a weight to the truth of the proposition: *the second document fulfils the IN of its time than the first does of its time* ${}_td_i$". Conversely if in $\nu_{i,j}$, ${}_td_i$ and ${}_td_j$ are about the same concepts while ${}_td_i$ is more relevant then the value of $|\nu_{i,j}|$, according to **only** these two document, gives a weight to the truth of the proposition:*the first document better fulfils the IN of the its time than the second does of its time*. Therefore $\nu_{i,j}$ is a way to decide between documents. The case where there are no concepts in common between the two state sets, ν_0 is set to ∞ indicating that the two paths are *incomparable*[65].

Applying this interpretation over all $\nu_{i,j}$ on the numerator of ν_0 in Equation 4.10 gives rise to the next proposition, if $\sum_{i \in {}_t\chi, j \in {}_t\chi'} \nu_{i,j}$ is positive it gives a weight to the truth of the

---

(naturally) combines the dynamic and static information which pertain to differing semantics the static corresponds to the concept of similarity from the vector space model where the dynamic is a probability, the combination of which indicates an interesting semantic (as addressed in Section 4.3.4.7). Secondly, in the way it keeps the two types of information separate in its representation (in a natural way) thereby facilitating a type of modelling (using the utility function) that preserves the categorical difference in the meanings of the information as the static corresponds to a collection property whereas the dynamic corresponds to a user property.

[64] If a document has higher relevance value then it is more needed, and it is said that the concepts in that document are more *needed in that document than in another document with the same concepts but lower relevance value*.

[65] This indicates the case when the documents in either of the path do not have any words in common, such that the inner product between any document in one path with any document in the other path is zero.

following proposition *the first state better fulfils the IN of its time than the second does in its time*, and if it is negative it gives a weight to the proposition *the second state better fulfils the IN of its time than the first does in its time*.

In a way it is as if each $\nu_{i,j}$ *voted* for a certain state with the numerator of ν_0 thereby indicating the overall winner and margin of win. Interpretation is aided by analogising the above with the idea of potential difference in electronic circuits. Potential difference (PD) is usually used to describe the difference between a initial and final point of an electrostatic potential. Between two points a PD gives rise to a 'force' called an electromotive force or EMF that tends to push electrons or other charge-carriers from one point to the other. The PD between two charges is a product of the electric field between them and their distance. One can regard each $\nu_{i,j}$ as a PD as shown in Fig. 4.5. The value of the numerator of ν_0 can be interpreted as the resultant potential difference between two sets of points on an electric circuit. Realising that the denominator of ν_0 is dividing up the EMF, it could be interpreted as electrical resistance, making ν_0 the current: the rate of flow of electrons between the two sets of points. Interpreting ν_0 as flow of electrons is analogous to looking at the summands of ν_0 in Equation 4.10 as a *flow of votes*, each vote indicating the superiority of the respective document or term, in fulfilling the information need of its time.

Special Cases Discussion in the previous section corresponds to the overwhelming majority of search paths that could be generated by a retrieval systems following the OM. There are however two families of special cases, both referring to patterns in the differences in relevance, where Equation 4.10 does not apply. The first is the case where all the relevances of documents in both states are equal giving $\nu_0 = 0$ which cannot be interpreted in the way discussed previously. In this case, the relevances cannot be used to discern between the states and the best one can do is use the information about the similarity between documents from the inner product in Equation 4.10 which collectively provide a measure for the average cosine similarity between documents in these states, Equation 4.11. In the following equations $|\psi|_d$ is the number of document terms in ψ, recall that this represents the number of documents with which the user can interact at time t.

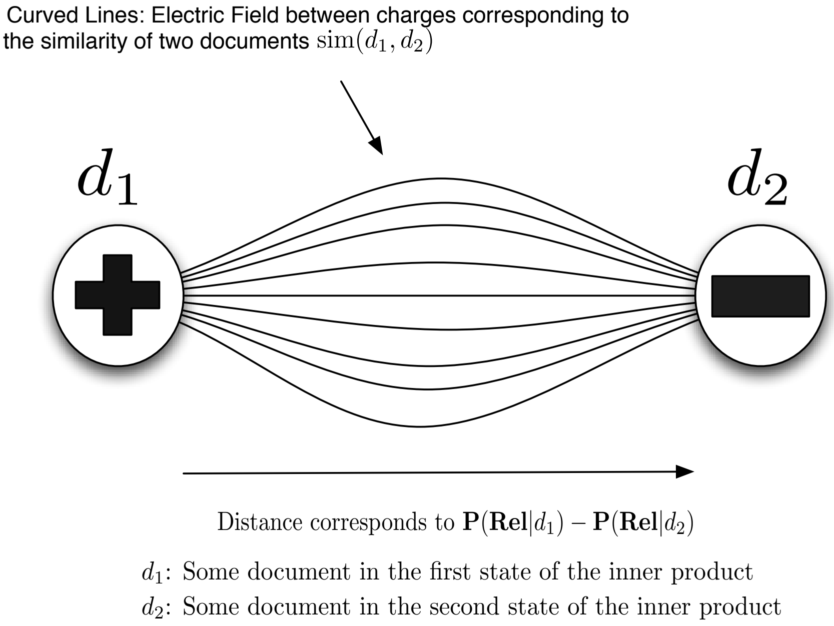


Figure 4.5. Inner product as Potential Difference: An analogy to aid interpretation of inner product between states

$$\begin{aligned} \text{If } \left(\mathbf{Pr}\left({}_t d_i\right)\right) &= \left(\mathbf{Pr}\left({}_t d_j\right)\right) \\ \text{and } \left|\left(\mathbf{Pr}\left({}_t d_i\right)\right)\right| &= \left|\left(\mathbf{Pr}\left({}_t d_j\right)\right)\right| = 1 \\ \text{Define measure for comparing states as:} \\ \nu_1({}_t\psi, {}_t\hat{\psi}) = Re\Big(\Big\langle {}_t\psi \Big| {}_t\widehat{\psi} \Big\rangle\Big) &= \frac{\sum\limits_{i \in {}_t\chi, j \in {}_t\chi'} \langle {}_t d_i | {}_t d_j \rangle}{\left|{}_t\psi\right|_d \left|{}_t\widehat{\psi}\right|_d}, \\ 0 \le \nu_1 &\le 1 \end{aligned} \tag{4.11}$$

The other case is when the relevances in the first search/state are all equal and equivalently for the second state but unlike the above case the two relevances are unequal. This means that all the differences between the probabilities of relevance of documents in Equation 4.10 are equal, say α, thus $\nu_0 = \alpha$ as the term $\sum\limits_{i \in {}_t\chi, j \in {}_t\chi'} \langle {}_t d_i | {}_t d_j \rangle$ vanishes. The remaining difference has no relation to information contained in the corresponding document thus Equation 4.12 is proposed, which is simply the common difference multiplied by average cosine similarity.

$$
\begin{aligned}
&\text{If } (\mathbf{Pr}\,({}_t d_i)) \neq (\mathbf{Pr}\,({}_t d_j)) \\
&\text{and } |(\mathbf{Pr}\,({}_t d_i))| = |(\mathbf{Pr}\,({}_t d_j))| = 1 \\
&\text{Define measure for comparing states as:} \\
&\nu_2({}_t\psi, {}_t\hat{\psi}) = \alpha Re(\left\langle {}_t\psi \middle| {}_t\widehat{\psi} \right\rangle) = \alpha \frac{\sum\limits_{i \in {}_t\chi, j \in {}_t\chi'} \langle {}_t d_i | {}_t d_j \rangle}{|{}_t\psi|_d \left|{}_t\widehat{\psi}\right|_d}, \\
&-1 \leq \nu_2 \leq 1
\end{aligned}
\tag{4.12}
$$

The ν_2 measure can also be interpreted in the way ν_0 was in preceding discussion.

4.3.4.8 Discussion of Utility Measures

The prior section gives a physical interpretation of the inner product between states as a *potential difference*. As the utility function is related to correlation measures, in particular, measures that correlate rankings of documents or of a set of words with another set of words such as topic correlation in [WJ04], utility can be interpreted as a correlation measure between states, search paths and/or documents. In this respect it is more similar to the utility function from Prospect Theory [KT79] than the measures in [WJ04]. However, as the utility measure is a signed measure it is not comparable except semantically with such measures as in [WJ04], further it is not testable with the traditional collection data unless individual retrieval scores are provided for the documents resulting from a query.

State change can be interpreted as an uncertain inference which `AU` uses to decide if and how to change state. In a cognitive sense, the method used to calculate the uncertainty in a proposition $\psi_1 \xrightarrow{\text{evolve}} \psi_2$ corresponds to ones personality. For example, for a real user, potential future states are considered and valued, and whether the user wants to change state to ψ_2 in the future depends on the value system. One important point here is that in analysing the utility measure it is apparent that the dynamic search information recorded by probabilities which contain information about user interests is naturally combined with information about the corpus (tf-idf weights). The ‘personality’ of the utility function is characterised by the fact that it *knows* by the probabilities of relevance, what the user interests are, and *then it inclines towards the topics which better pertain to this interest*, normalising this inclination according to similarity between the topics (as in Equation 4.10).

Document Level Utility Function States represented as superposition of documents in an orthonormal document basis only possess dynamic weights and the utility (or inner product) between two states represented in this way (without re-representing in term basis) is zero as:

$$\begin{aligned} {}^{d}\nu(\psi_a, \psi_b) &= \sum_{d \in \psi_a, d' \in \psi_b} \mathbf{Pr}(Rel|d) - \mathbf{Pr}(Rel|d') \\ &= \left(\sum_{d \in \psi_a} \mathbf{Pr}(Rel|d) \right) - \left(\sum_{d' \in \psi_b} \mathbf{Pr}(Rel|d') \right) \end{aligned}$$

As the probabilities of relevance for documents in each state sum to 1:

$$ {}^{d}\nu(\psi_a, \psi_b) = 1 - 1 = 0 \qquad (4.13)$$

The document level utility function is thus named to indicate that the states it takes as parameters can be symbolically written as a superposition of documents; it assumes that there is enough information about documents to allow them to be represented in a term basis. The utility at the document level between states, meaning that the state is represented with a bases consisting of documents, denoted ${}^{d}\nu(\psi_1, \psi_2)$ *which is calculated following the re-representation of documents in term basis* combines static term weight with dynamic document weights (probabilities)[66]. The utility in this case indicates if the first state or the second is better and facilitates the definition of the information need uncertainty principle in the proceeding that is semantically related to the logical uncertainty principle [vR00][67].

Information Need Uncertainty Principle The direction of *energy flow* between two information items (which include states) is ascertained according to the[68] *Information*

---

[66] Therefore document-level utility is different from term-level utility as the former contains both term-level static and dynamic weights where as the former has one of either type.

[67] It should be emphasised that the relation here is semantic at this point and requires further study in a logical framework to determine if the link is beyond the conceptual level.

[68] Note that as described in Chapter 3, IN exists between any two states where a state could be that as defined in this chapter or the state of a search component, further details of state are given in the next chapter.

Need Uncertainty Principle: If both ${}_t\psi_1$ and ${}_t\psi_2$ are possible future states for the same point in time and $-1 < {}^d\nu({}_t\psi_1, {}_t\psi_2) < 0$ then ${}_t\psi_1 \xrightarrow{\text{‘energy flow’}} {}_t\psi_2$ which means that ${}_t\psi_2$ contains more relevant knowledge relative to the system's concept of user interests at that time. In summary, one could say that given any two states ${}_t\psi_1$ and ${}_t\psi_2$, a measure of the uncertainty of ${}_t\psi_1 \xrightarrow{\text{‘energy flow’}} {}_t\psi_2$ related to a given document space $\mathcal{H}$, containing static (term) weights and dynamic weights (current probabilities relevance), is determined by $\nu\left({}_t\psi_1, {}_t\psi_2\right)$. The document-level utility applies to this uncertainty principle on the assumption that only document-level dynamic information (probabilities of relevance) are available, if instead term level dynamic information is also available then the term-level utility function can be used. Hence, in addition to being a measure of uncertainty, the utility measure can be interpreted as a measure for the (potential) information flow between states which indicates the potential to change to another state and the usefulness (in terms of static and dynamic weights) of this change between two states of search. Further, it can be used as rank correlation measure[69], except, due to including dynamic information instead of purely static information as in usual correlation measures (see [WJ04] and [CMZ05]), it would be difficult to test; Section 6.6 comments further in this regard. The utility is of more use when employed in exhaustive fashion for analysis of collections in the way elaborated by Section 6.4.

4.3.5 Measures over States

A simpler representation of state is to set the argument of the complex coefficient of term weights to the probability of relevance such that successive applications of change operators to a state represented in term basis gives the document representation at time $\tau + 1$ of the form:

$$\tau > 1, {}_{\tau+1}D_i = \left(w_{i,j}\right) {}_{\tau}e^{\jmath\left[\mathbf{Pr}_0(Rel|w_j)+\sum_{t=1}^{\tau-1}\delta\mathbf{Pr}_t(Rel|w_j)\right]}|w_j\rangle\langle w_j| \tag{4.14}$$

where $\delta\mathbf{Pr}_t(Rel|w_j)$ is the change in probability of relevance of term j at time t and $0 \leq \mathbf{Pr}_0(Rel|w_j) + \sum_{t=1}^{\tau-1}\delta\mathbf{Pr}_t(Rel|w_j)] \leq 1$. Measures over documents that employ static information have forms such as $f_{\text{stat}} = f_{\text{stat}}\left(|d_i|\right)$, ${}^{d_i}f_{\text{stat}} = f\left([w_{i,j}]\right)$, or ${}^{d_i}f_{\text{stat}} =$

---

[69] Since it compares two sets of documents (states) including their dynamics/probabilities of relevance, thereby using more information than a simple rank.

$f\left(\left[w_{i,j}\right], \left|d_i\right|\right)$ or more generally[70]:

$$f_{\mathsf{stat}} = f_{\mathsf{stat}}\left(\left[f_{\mathsf{stat}_i}\left(w_{i,j}\right)\right], f_{\mathsf{stat}_j}\left(\left|d_i\right|\right)\right) \tag{4.15}$$

4.3.5.1 Mixed Static/Dynamic Measures

Measures that consider the dynamic information would be would then be functions of the type[71]:

$$f_{\mathsf{dyn}} = f\left(\arg\left(d_i\right)\right) \text{ or } f_{\mathsf{dyn}} = f\left(\left[\arg\left(w_{i,j}e^{\jmath\theta_j}\right)\right]\right) \tag{4.16}$$

or more generally:

$$f_{\mathsf{dyn}} = f\left(\left[\arg\left(w_{i,j}e^{\jmath\theta_j}\right)\right], \arg\left(d_i\right)\right) \tag{4.17}$$

In the case that both dynamic and static information is to be used the measures are functions of form $f_{\mathsf{mix}} = f(f_{\mathsf{stat}}, f_{\mathsf{dyn}})$ or $f_{\mathsf{mix}} = f\left(\left[w_{i,j}\right], |d_i|, \left[\arg\left(w_{i,j}e^{\jmath\theta_j}\right)\right], \arg\left(d_i\right)\right)$, a simple sub-form is $f_{\mathsf{mix}} = f\left(\theta_j * w_{i,j}\right)$ and example of which is $f_{\mathsf{mix}} = \frac{1}{|T|}\sum_j w_{i,j}\theta_j$[72]. A quantitative example of term-level utility is given below:

Example 1. *[Term-level Utility] Let the relevance values for a 4 term collection be* $[0.2, 0.1, 0.3, 0.4]$ *then if* ${}_0D = w_1|t_1\rangle\langle t_1| + w_2|t_2\rangle\langle t_2| + w_3|t_3\rangle\langle t_3| + w_4|t_4\rangle\langle t_4|$ *and* ${}_1U$ *is the operator that sets up the document with a probabilities of relevance for terms. The utility of this document would be* $\nu({}_1D) = w_1(0.2)+w_2(0.1)+w_3(0.3)+w_4(0.4)$ *where each weight* w_j *quantitatively specifies the extent to which term* j *is represented by this document and the probability denotes to what extent this matters to the current interpretation of IN.*

4.3.5.2 Measures as Functions of Other Measures

Measures (static or dynamic types) over a state can be functions of measures over documents $f\left(\left[{}^{\mathsf{mix}}f_i\right]\right)$ or functions of term coefficients $f([w_{i,j}], \theta_j)$ when states are represented in term basis, for example $\psi = \sum_i d_i = \sum_j \left[w_{1,j} + w_{2,j} + \ldots w_{|d|,j}\right]e^{\jmath\theta_j}$. The utility mea-

---

[70] Recall that square brackets denote a list, thus this function is *a function of a list of functions of term weights, paired with a list of functions of document-level weights*, these weights are attributed to documents as a whole and not terms.

[71] In the case a document contains all terms its argument is the sum of the probabilities of relevance for all terms which in a term basis is a sum to 1.

[72] The sum here is normalized by $|T|$, the number of terms.

sures of the prior sections are mixed dynamic and static measures over two states and are of the form $f_{1\psi,\ 2\psi} = f\left(\left[{}^{\text{mix}}_{t_1} f_i\right], \left[{}^{\text{mix}}_{t_2} f_i\right]\right)$. At a higher level, the measures is over multiple states, which can be used to represent states of the system at a set of different times, or more generally multiple sets of states which represent groups of states of search at different time periods, $f_{[\psi]} = f\left(\left[\left[{}_{\tau_1}\psi\right], \left[{}_{\tau_2}\psi\right], \ldots\right]\right)$. The natural form of functions on the vector space is through inner products. If documents are represented in term basis, the inner product between documents means document similarity, between states it means average document similarity (upon normalisation) in the collection, it is therefore a static measure as the complex parts vanish on multiplication of document vectors. If instead one uses a document basis and projects onto a term basis (as in previous section) then the inner product can be used to create mixed static and dynamic measures. It is the vector space structure of Hilbert space that allows such measures.

4.3.5.3 Trace Measures

Consider the matrix $\overline{{}_0D} = \left({}_0D\right)^{-1} = \left(|{}_0d\rangle\langle{}_0d|\right)^{-1}$ where $|{}_0d\rangle\langle{}_0d|$ is the initial diagonal matrix in term basis holding static weights of terms with respect to the document, where it is assumed that the dynamic weights are only added in $t = 1$ thus each weight (at $t = 0$) is a real value[73]. The matrix $\left(|{}_0d\rangle\langle{}_0d|\right)^{-1}$ consists of diagonal entries $\overline{{}_0D}_{j,j} = \frac{1}{\left(|{}_0d\rangle\langle{}_0d|\right)_{j,j}}$. The operation $D\left(\overline{{}_0D}\right)$ cancels the static weights of the arbitrary time document matrix $|D\rangle\langle D|$, and the operation:

$$\operatorname{dyn}(D) = \jmath \ln\left(D\overline{D}\right) = \sum_j \ln\left(e^{\jmath\theta_j}\right)|w_j\rangle\langle w_j| = \sum_j \mathbf{Pr}(Rel|w_j)|w_j\rangle\langle w_j| \qquad (4.18)$$

leaves the document matrix with dynamic weights (probabilities) on its diagonal such that $\operatorname{tr}(\operatorname{dyn}(D)) = 1$[74].

---

[73] This means that ${}_1D = U_1\left({}_0D\right)$ where U_1's diagonal entries are probability of relevances of terms.

[74] This resultant matrix is (in QT terminology) is a *density operator*. The density operator of a system are usually used to generate probabilities, whether this formulation can be used to create a quantum-operator based probabilistic framework is an open problem and is not further addressed in this thesis.

Term-level Utility The term-level utility measure of the prior section for a single document can then be expressed as:

$$^{w}\nu\left(d_i\right) = \mathrm{tr}\left({}_0D\ \mathrm{dyn}(d_i)\right) = \mathrm{tr}(\mathrm{dyn}(D)D) \tag{4.19}$$

where each entry in the resultant matrix inside the trace has (mixed) diagonal entries $w_{i,j}\mathbf{Pr}(Rel|w_j)$. Other utility functions can be envisaged in the form $\mathrm{tr}\left(T_1\ D\ T_2\ \mathrm{dyn}(D)\right)$ where T_1 and T_2 can be used to denote a particular method of combining the static and dynamic weights.

The *term-level utility over a single document* $^{w}\nu(D)$ is defined as:

$$\begin{aligned} ^{w}\nu(D) &= \mathrm{tr}\left({}_0D\mathrm{dyn}(D)\right) \\ &= \mathrm{tr}\left(\sum_j w_j\mathbf{Pr}(Rel|w_j)|w_j\rangle\langle w_j|\right) \\ &= \sum_j w_j\mathbf{Pr}(Rel|w_j) \end{aligned} \tag{4.20}$$

whereas the *term-level utility measure over a single state* $^{w}\nu(\psi)$ is defined as follows (assuming normalisation of weights):

$$\begin{aligned} ^{w}\nu(\psi) &= \sum_i \mathrm{tr}\left({}_0d_i\mathrm{dyn}(d_i)\right) \\ &= \mathrm{tr}\left(\sum_i {}_0d_i\mathrm{dyn}(d_i)\right) \\ &= \mathrm{tr}\left(\sum_i w_{i,j}\mathbf{Pr}(Rel|w_j)|w_j\rangle\langle w_j|\right) \\ &= \sum_j\sum_i w_{i,j}\mathbf{Pr}(Rel|w_j) \end{aligned} \tag{4.21}$$

where[75] the i^{th} summand gives a measure indicating the usefulness of that document in a search where term weights are dynamically modified (due to the probability of relevance changing) and take into consideration the extent to which relevant terms are represented in that document (the $w_{i,j}$).

---

[75] Note that the trace operator in Equation 4.21 applies to a matrix in term basis.

Mixed-Level Utility If one takes the document-level utility function for a state ${}^{d}\nu({}^{t}\psi_1, {}^{d}\psi_2)$ with the parameter states in different bases but without being re-represented into the same basis. Thus, the requirement of re-representing states from document to term basis for one of the states (in its parameter pair) before it is sent to the function, is removed. Such a function, denote by ${}^{m}\nu({}^{w}\psi_a, {}^{d}\psi_b)$, is said to be a *mixed-level utility* and is defined as follows[76]:

$$ {}^{m}\nu({}^{w}\psi_a, {}^{d}\psi_b) = \sum_{i_a, i_b} \mathrm{tr}\left({}^{w-basis}D_{i_a} T_{d\rightarrow w} \mathrm{dyn}\left({}^{d-basis}D_{i_b}\right)\right) \tag{4.22} $$

where $T_{d\rightarrow w}$ is the projection (re-representation) that denotes the change of basis from the document space (in document basis) to term space (where term vectors form an orthonormal basis).

The document basis in $\left({}^{D-basis}D_i\right)$ of Equation 4.22 is useful when document-level dynamic information is present in one state but term level dynamic information is not. Re-representing this in term basis introduces term-level static information to appear and allows document level dynamic and term level static information to be combined[77]. In this way, one can do $\nu(D_1, D_2)$ or $\nu(\psi_1, \psi_2)$ which is similar to the term level utility as it is a function of term level utilities of documents.

The value of $|\nu(D_1) - \nu(D_2)|$ or $|\nu(\psi_1) - \nu(\psi_2)|$ (i.e. regardless of basis) refers to the effectiveness of the document and states respectively in representing the information need of the time; whereas, the sign of the difference indicates which document/state is better at representing the IN of the time. The utility measure between two states or documents can also be seen as a linguistic measure, as, if there was an 'optimal' state or document ψ_o or D_o respectively such that $\forall i, \nu(D_o) > \nu(D_i)$ or $\forall i, \nu(\psi_o) > \nu(\psi_i)$ then it is said that *the document D_o or state ψ_o is such that the IN's at the times associated with D or ψ, as represented by the probabilities of relevance at the time, are most representative of the information in the documents (corpus) or states respectively.*

---

[76] As in the definitions of other utility functions, assume that the static weights are normalized.

[77] A suitable representation (not addressed in thesis) would be required for combining document level static and dynamic information onto term level static and dynamic information, as the dynamic information stored would require to be suitably combined.

4.3.5.4 Incorporating Context into Measures

The wave function approach gives rise to utility functions due to separate representation of dynamic and static information. The observation technique of using QT (R_m) in Section 4.3.3 gives experimental results[78] which are denoted by probabilities, this is as expected due to semantics of measurement. The wave function is a representation of phenomena, it has to refer to[79] all that is known, which for IR means the information about the corpus to start with and the interpretation of relevance over time. Whereas, for the observation technique one can store whatever is needed and one is not required to store all information in the representation. Clearly the wave-function style representation presented so far is incomplete as there are many other types of information pertaining to contextual factors that require to be stored. Thus, the prior models are limited due to this, but the observation method, which is required to 'ask questions' to a representation and formulate arbitrary contextual information (that needs to be stored for search), allows arbitrary information to be 'stored by phenomena' as it does not store all this or require to model phenomena that stores it. Instead, it simply 'records' the effects of phenomena pertaining to these contextual factors. Initial work addressing the modelling of contextual factors by modelling them as observables, referred to as the measuring method in Section 4.2.3.3, can be found in [MW07b] and [MW07a].

4.3.6 Change of State and Dynamics

As demonstrated in prior sections QT allows structuring of state change which corresponds to structuring of relevance feedback. A relevance feedback structure needs to consider user interactions and context, QT formalism provides several theoretical devices with which to represent this. The problem in IR is that there is little ability to structure feedbacks, this is not only a problem of not having a unified model for feedback and relevance value updating[80], but also due to the fact that the user and system are not treated as separate in

---

[78] Recall that R_m makes observations through *measurement experiments*.

[79] The need to store or refer to all known quantities is a semantic requirement that is at most a weak one due to our *customised* use of QT, the following chapter elaborates further.

[80] Vector space model requires whatever feedback to be mapped onto vector rotations, thus a lot of information has to be diluted into one vector.

the feedback structure as the user is not modelled, it their interactions which are 'recorded'. The perspective here is that of a third entity's view of 'relevance', it is a confusing point of view[81]. One way to separate this, and do structured user modelling is to presume what relevance and relevance feedback is, and how it changes. The next section presents an initial approach to adding structure to relevance feedback, and in turn presents a method for simulating searches with dynamically changing information need represented by changing probabilities of relevance.

4.3.7 Simulation Design

In the vector space model over $\mathbb{C}$ the corpus can be represented by a superposition of document vectors each of which is further represented $D_i = \sum_j^T w_{i,j} e^{\jmath\theta_j} \left| t_j \right\rangle$ where the argument of the complex numbered weight $\theta_{i,j} = \theta_j$ denoting that this part of the weight applies to the j^{th} term in all documents. The $w_{i,j}$ is a static relation between the (semantically) lower level terms and the higher level documents, in terms of language modelling the terms $w_{i,j} = f(\mathrm{tf}_{i,j}, \mathrm{idf}_{i,j})$. Dynamic search information such as probabilities of relevance, would be encoded in θ_j as done in the prior sections. Thus the complex field in the Hilbert space allows one to store these two (i.e. static and dynamic) important types of information about the state of a term. If the field had even more parameters then more information related to the context of the term and specifically linguistic characteristics could be stored. Currently, these other properties would require to be stored by alternate representations of the term space by means of the measurement method (see Section 4.3.3 and Section 4.3.5.4)[82]. Modifying the real part of the term weight would require that the document vector is multiplied by a vector $v \in \mathbb{R}^{n\times 1}, \arg\left(v_i\right) = \pi$.

It is convenient to place dynamic search information as the argument of the complex weight, as change in the search process can then be recorded in terms of *word states* and upon interaction or change to the corpus the arguments of all complex term weights could be changed independently of the static part of the weight[83]. This would be done by multiplying

---

[81] Is it the system's, user's or researcher's point of view of relevance that is being modelled when in a particular state of search?.

[82] Thus a context state would need to be specified as a superposition (or other association) of term spaces, or a term state would require to be specified in terms of lower-level sub-states.

[83] The static and dynamic information could have been represented as one number but then one would

the diagonal document matrix $D_i = \Sigma w_{i,j} e^{\jmath\theta_j}$ by a unitary matrix ${}_tU = \Sigma_{i=0}^{|T|} e^{\jmath\delta\mathbf{Pr}_i(Rel|w_j)}$ where $\delta\mathbf{Pr}_i$ represents the individual change of dynamic value (of relevance in this case) in term i.

4.3.7.1 Exploiting Symmetry

Assuming that the initial probabilities of relevance of the terms $\mathbf{Pr}(Rel|w_j)$, and also changes in relevance at each time during a search $\delta\mathbf{Pr}(Rel|w_j)$, one can define a set $[{}_tU], 0 \leq t < T$ which is the change matrix representing the changes in relevant at time t. If only one term's weight is to be changed then ${}_tU$ is defined such that the arguments of all weights $\arg({}_tU_{i,i})$ are nil except one. The dynamics of a search in terms of higher-level semantics can be denoted by representing a document by a set of these semantics, such as latent semantics through performing linguistic analysis in terms of matrix operations, and in this case each change matrix would be defined in a basis of semantics as opposed to a term or document basis. Assuming that one has the probability of relevance updates (from t_1 to t_2), define groups such that for time periods $t_1 \leq t < t_2, [U_i] \longrightarrow G\left([I, [U_i]], \times_{nxn}\right)$ with identity I where the group operation is matrix multiplication and n is the dimension of the semantic. The n denotes the number of terms, if change analysis is in terms of words, or it is the number of semantic dimensions if modelling latent semantic analysis.

Commutativity The group is Abelian if $({}_1U)({}_2U) = ({}_2U)({}_1U)$ as this corresponds to the commutativity of addition since applying ${}_tU$ to terms implies addition/subtraction of the arguments of the complex numbered entries. The binary operation on G is an associative operation, it also contains an identity thus if it contains inverses (inverse changes of relevance) then it is a closed group. Thus it can be said that a search operation whose dynamic behaviour in terms of changes of probabilities of relevance, that can be denoted by a closed group, is a *neutral or 'conserved' search* as the system's concept of the user interests at the end of the search is equal to that at the beginning. As if each change has an inverse within the search then this implies the interest in that particular term is 'taken back' or 'reset', if this is true for all terms over the course of a search then user interests stay invariant over a search, as seen by this system design. More realistically, it could be

---

require to store details about how to separate the weights for independent static/dynamic change analysis.

said that a subgroup ${}_{t_1,t_2}X \subset U$[84] of changes from t_1 to t_2 were such that the user interests were invariant over this time, that the user made some decisions but later made opposing decisions and their interests changed back to what they were at t_1. Alternatively one could define a set of change matrices and corresponding groups: ${}_tU_{a,b,c} = [[{}_tU_a], [{}_tU_b], [{}_tU_c]]$ that denote the changes in term number ab, and c respectively over the course of the search, and more generally one could let ${}_{[[t_1,t_4],[t_5,t_8],\ldots[t_{10},t_{11}]]}U_{a,b,c}$ denote changes over sets of terms or sets of semantics in three or more periods during a search.

In general ${}_tU$ is not a closed group as not all of its elements may have inverses. However, when a subgroup is closed there are interesting implications for user interaction which can be related to the *conservation of information need* relative to a set of terms over a period of time as the IN or user interest, for those terms has changed then changed back to what it originally was at the start of the time period. One can also discuss changes in terms of a cyclic group ${}_t\dot{U}$ where all changes in the group are in terms of one particulars group ${}_tU = [I, U, U^2, U^3, \ldots U^n]$, and since each U has complex values each with an encoded probability the group elements represent the probabilities $0, k, 2k \ldots (n-1)k, nk$ where $0 \leq k \leq 1, nk \leq 1$[85]. A cyclic group design for changes for arbitrary *k* values would denote a structured dynamics of changes for term relevance, semantic/concept relevance or document relevance (once the static is combined with the dynamic part of the complex weight).

A simulation design requires specification of how the user interests would change over time. Representing changes of relevance in terms of groups seems an organised way to develop these dynamics. The simplest simulation designs are closed subgroups which allow us to map the concepts like '..user is initially interested in topic X until he sees topic Y by which time he is no longer interested in X but in Z' to a group that includes inverses of all its elements[86]. Thus simulation specifications be compared on a conceptual and mathematical level. On the other end real-life user experiments that provide a set of

---

[84] This subgroup could denote a set of changes that follow a semantic, for example it could denote a set of changes in relevance that resulted from a particular type of user behaviour as recorded by the interface.

[85] As the maximum change in a probability has magnitude 1.

[86] Note that the idea of a user no longer interested in a topic should be taken to mean an inverse *as if* they were not interested in the topic in the first place. Although this is an unrealistic depiction of change of user interests it is of initial use until one can model a type of 'uncertain' or probabilistic, inverse or disinterest, in a topic.

$\mathbf{Pr}(Rel|d)$ over time can be approximated by predefined groups denoting dynamics and partially constituting user models[87]. Complex user behaviours albeit all based around the variable of probability of relevance[88] can be developed for simulation and for modelling complex real-life user behaviours.

4.3.7.2 Morphisms

A morphism between two groups or subgroups of changes would denote the user changing strategy, if $\phi : {}_tX \subset {}_tU :\mapsto {}_tY \subset {}_tU$ and $X = Y$ then this ϕ is an *automorphism*, a trivial example is a permutation on the group ${}_tU$ that can be taken to denote an alternative ordering of user interactions which can be used to suggest future behaviours and subsequent changes to relevance values[89]. For a particular simulation design, one can investigate the effect of the order of user interactions by considering all or some permutation based automorphisms of that group, a randomised order of user interactions would also be represented by the permutation of a group of pre-defined possible actions, this is one way to address random user behaviour[90] and a way to pick possible future states for a search. If $X \neq Y$ then ϕ is a homeomorphism, a map between two groups that preserves the groups properties. This can be used to denote a change of strategy and therefore can be used to semantically classify sets of simulated user strategies. For example, the homeomorphism ${}_t\dot{U}_{k=-0.1} \longrightarrow {}_t\dot{U}_{k=0.2}$ denotes an *decelerated decrease in the IN change for a set of terms* whereas in general one can have a set of homeomorphisms or strategy changes like ${}_t\dot{U}_k \longrightarrow {}_t\dot{U}_{f(k)}$. Ageing models from ostensive retrieval [Cam00] can be represented by letting $\arg\left({}_tU_{i,j}\right) = f(t)\mathbf{Pr}(Rel|w_j)$[91] or more generally by functions $\zeta : (t, U) \mapsto U$ that modify the probability accordingly. It is proposed that such discrete structuring methods are important for modelling the user's information need change.

---

[87] This is where approximate groups would help to add uncertainty thus allowing us to say that the user behaves approximately like G_1 in a period of time and like G_2 at a later period, where each of group corresponds to a particular behavioural semantic.

[88] See next two chapters for discussion pertaining to introducing other variables.

[89] Can also randomly permute interactions for the purpose of a randomised analysis of search behaviour.

[90] This is semantically similar to latin square design for statistical experiments in traditional IR evaluation.

[91] $f(t)$ is an ageing function.

4.4 Summary

This chapter suggested techniques for representing search phenomena, and discussed especially the representation of the dynamics of search. The next chapter extends these techniques by introducing a more general representation that need not conform to the mathematical rules of a geometrical space, by which it proposes to simplify the modelling of search phenomena and accommodates a format for semiformal discussion that is more suited to addressing the definition problem in IR research than a purely QT language. The QT framework shows that one can have a high-level operator theory which can be represented by a 'lower level' theory such as matrix theory, and itself be at a 'lower level' relative to physics semantics like energy. The framework shows that these different levels of representation can work together. Also a level higher than that of operators are the concept of groups of operators as in Section 4.3.7. Thus group theory can be used to discuss physical events in terms of their 'group' properties and then find a theory, like matrix theory, where one can represent it so as to be able to derive quantitative understandings of geometrically, probabilistically and/or logically[92] orientated concepts.

Using $\mathcal{H}$ and corresponding ways of using $\mathcal{H}$ from QT is useful for modelling specific search phenomena yet there is much to use from QT[93]. The first and second postulates of quantum mechanics say that the QT provides a representation for states and their change but cannot tell us what they ought to be as that would require concepts related to physical reality to be adapted for an abstract process like search. In this chapter only the first two postulates of QT were used for modelling search (Section 4.3.4) whereas measurement observables (third postulate) and composite systems (fourth postulate) were not. The semiformal language in the next chapter is such that it includes the third postulate on a semantic level, in that, the system changes with measurement, and it also includes the fourth postulate such that composite systems are represented as a 'semantic union' of states (which when represented in QT correspond to composition and product spaces[94]).

This chapter showed that using QT as a symbolic language can provide a useful notation for

---

[92] Logical aspects were not covered in this chapter, see [vR04], [Gri02] and [BC81] respectively for expositions of logic in QT in increasing detail.

[93] And possibly its related disciplines.

[94] Tensor products of Hilbert spaces.

representing IR concepts. The physics semantics in imported by use of the QT framework are also useful. For example, represented in QT form, the state of the system represented as a superposition of sub-states corresponds to the representation of some part of the user as a potentiality or concepts [AG02]. The similarity in the representations of user and system components, state of mind and search path respectively suggests that perhaps it is useful from an IR modelling point of view to perceive them as similar, as suggested by the assumption that HHIR is the optimal IR. In this way, QT representation allows new insights about the similarities between user and system modelling, in particular those similarities revealed by symmetry in representation. The next chapter elaborates on this modelling of the system and user, and discusses the symmetry in such modelling.

One difficulty with modelling as highlighted by our example state representation in terms of superpositions pertains to the question: which observers view should be modelled? Should it be that of a researcher looking at search, if so should it be for a particular search process or for all search processes? As explained in the previous chapter, the observer is the researcher and it is their view that should be modelled, the next chapter clarifies this further in terms of new notation.

—Actions differ because the inspirations of the states of being differ.

Ibn Ata'illah al-Iskandari (∼1250-1309)

5

On Search as a Communication Process

5.1 Introduction

In Chapter 3, a very general epistemology and ontology of a search process was given. The concepts of a search state, statements associated with states, agents participating in a search (including the researchers) were discussed. In the previous chapter quantum theory was introduced to show how these same notions are employed in a science to articulate the physical world, a science which has had empirical success and is very useful for theoretical analysis, perhaps more so than any other modern theory.

This chapter builds a framework for describing and analysing search processes borrowing from the techniques that QT employs to build its own theoretical framework to abstract the physical world for analysis. The framework developed is not a quantum theory for search but a *pre-quantum theory* in the sense that it abstracts the search process simplifying the process of mapping some aspect of a search, if beneficial, to a formal theory like QT. The framework is also one which is a *pre-computational* nature in that it abstracts the search process simplifying the process of mapping search concepts to algorithms for implementation in software or for computational analysis. It is therefore a *middle language or a middle framework* that, as shall be argued, is a necessary medium to join the requirements of having a formal framework where search can be theoretically analysed and the findings implemented in search models. Several variables make this task difficult, including the four key problems in Section 2.6, the framework presented does not aim to be an end in itself to all these issues but a foundation that breaks down these problems into hopefully easier tasks that can be approached independently.

This chapter starts with discussion of a search state and its particulars. It should be noted that the representation of search states which hold static information about the search, and flows that hold the dynamics of search, by which I mean the notation used to refer to these elements and the way in which the notations are associated to one another, denote in themselves the majority of our framework. In the same way grammar and logic of a language is embedded into the use of the language and therefore can be extracted from these texts, the notation and application of notation to describe search phenomena[1] indicate the novel 'grammar for IR research' that is being proposed. Initially

---

[1] This is addressed in more detail in Chapter 6.

I elaborate on the visualisation of states as introduced in Chapter 3. Visualisation is an important aspect as if the middle language proposed accommodates derivation of software algorithms, formal models (of a QT nature), and also cognitive models, then ideally it should also accommodate diagramming techniques that are usually employed in each of these areas so that, at least initially, one is presented with familiar techniques for using our framework.

5.2 Visualisation in terms of stacks

There are paradigmatic differences between QT and IR in their operational methods and semantics (as briefed in Chapter 3 and Chapter 4). It is clear that the mapping to QT has mathematical benefits, with features like complex numbers and algebraic structures on the Hilbert space but it is unclear as to how one can generally use the QT formalism, what QT concepts cannot be used and more importantly what some QT concepts corresponding to the formalism (such as algebraic structure) mean in IR terms. A bottom-up approach would be to assess the benefits of each feature of the QT framework individually, however it is difficult to deduce IR meanings for IR models created in a QT formalism. Instead, as shown in the ostensive chapter, a 'top-down' approach is attempted with the premise that the apparent paradigmatic differences between IR and QT can be reduced if one no longer thinks of a search process in the traditional sense, according to the laboratory model, but instead in the way QT would perceive a process, as a physical process. In light of the previous chapter it is clear that the search process was abstracted in the ostensive chapter as a physical process in which there are a set of interactions between at least two physical systems. The diagrams Fig. 3.3 and Fig. 3.2 in Chapter 3 illustrating statements within agents are stacks, if further generalised so that each set of statements associated to a state are collectively labelled according to their use then one gets a diagram like the stack in Fig. 5.1. This visualisation, inspired by a computational perspective, is visually equivalent to a traditional network architecture diagram illustrating the design of the protocol of communication between two agents. It corresponds to a design method for visualising arbitrary search scenario designs. Stacks are labelled according to their purpose in a search scenario.

An investigation focusing on document/matching models which is at the memory/reasoning level may not need to discern between the gestures and physical layers, combining them instead. Visualisation in stacks implies a method of design unlike the laboratory model, since it freely includes (by design) other aspects of the user and system, including as in Fig. 5.1, the interface (gestures layer), hardware (physical layer), search strategy (session layer) and the corresponding 'components' of an user model. In terms of the stack in Fig. 5.1 one can say that a user interacting with a system involves a reasoning sub-process instigated by the memory layer, which then influences a search strategy (session layer) and activates corresponding gestures expressed by the physical layer representing their physical expression tools. The system in Fig. 5.1 is setup with the same design; it observes the user's physical action, interpreting a gesture in the context of other factors (such as prior user feedback) and updating its notion of user interests (memory) according to a user-interest update policy (reasoning). Memory and reasoning layers in the stack correspond respectively to the document and document model in the laboratory model. A search process consists of a set of sub-processes of different types, an *observational process* involving two or more separate agents, an *expression* and *interpretation processes* that happen within agents involving components; these are common principles for a stack based model of a search process. The sub-process of type interpretation denotes a general set of activities directed towards[2] the memory level, it corresponds to one set of changes that are internal to an agent upon another agent interacting with it. The other set of changes internal to an agent are those that lead it to react to the prior interaction, these changes are of the expression type and generally originate in the memory layer eventually influencing the agent's expressive faculty, the interface, denoted by the physical layer.

The stack design visually represents the shift in the traditional conceptualisation of IR toward the state-based conceptualisation as introduced in the ostensive chapter and detailed for physical systems in QT in the previous chapter. It was inspired by the QT way of abstracting physical sub-systems and their relationships. In terms of the physical semantics of QT, each layer is a physical sub-system and can only directly influence adjacent sub-systems which preserve the order inherent in search systems. The stack model does not presume any specification language for any layers. A clearer relationship between the user,

---

[2] This refers to the direction of influence and activity, thus for the interpretation sub-process necessitates that activity in lower layers precedes that in higher layers.

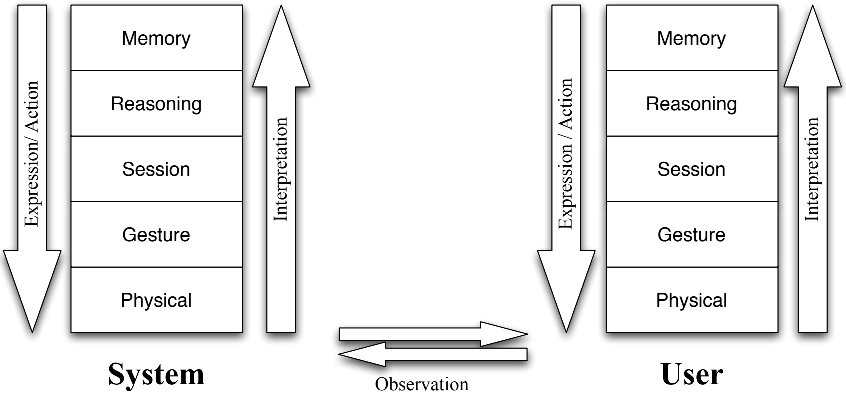


Figure 5.1. A Stack Model

definition and evaluation problems can be observed when these problems are visualised by stacks. First, as the research in [vR04] shows, the mathematical formalism of QT can be used as a language for modelling document and matching models, corresponding to the memory layer and reasoning layers. However, a QT specification of the reasoning in a search engine is theoretically limited as one cannot predict the relevance feedback, which would require models of cognitive behaviour in a way similar to why QT still requires physics and physical semantics (i.e. physical laws) in order to predict the behaviour of systems[3]. Second, as per the definition problem, retrieval research provides no general frameworks for modelling useful interfaces and interactions, thus the gestures and session (containing search strategies) layers have no general[4] modelling language. Similarly, on the user side, there are no general modelling languages for any of the layers in the stacks. Instead the literature of IR and related disciplines provide several models specified in different languages for each layer of each agent and no analytical way to compare between them. The methodology required for modelling corresponding to the visual stack diagram, a formalism for specifying each layer, inter-layer communication and between agent communication is presented in this chapter which aims to address the definition problem now that the conceptual problem has been resolved in Chapter 3.

---

[3] See Appendix C for a brief further discussion.

[4] There are no general accepted languages or techniques for describing user actions in an arbitrary interface. A specific example is in [Xie02].

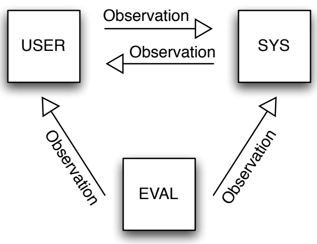


Figure 5.2. Model of an Evaluation

5.2.1 Modelling of Diverse Information Seeking Scenarios

The stack diagram is also used to depict the evaluating agent that could be a researcher[5], or if the search scenario is a simulation the evaluating agent denotes some part of the simulation software, and hence the evaluation process can be illustrated as a three entity process as in Fig. 5.2. The practice of modelling the measurement device in QT inspired this way of using stacks. An information seeking process, one which considers the broader experience of the user in the real life search process can involve several more agents some of which could be human agents. An example would be the case of a user interacting with librarians, automated search systems and other agents in a library. The information seeking process is visualised by adding more stacks to the design of the typical information retrieval model of two stacks. A distributed search scenario such as in peer-to-peer searching or searching through multiple indexes representing an intranet, can also be visually depicted in the same way by a set of stacks.

5.2.2 Topology of visualisation

States are structured where each substate corresponds to a sub-part or sub-system. Substates are related to one another according to whether they influence each other in a search. There are no intrinsic lowest elements in the set of all substates so a state has substates that may decompose into boundless number of further substates. There are extrinsic limits to substates of states such as that of the document state which is composed of term states that have no further substates[6], thus limits of the structure for states are

---

[5] As per the conceptualisation of agents and ostension in Chapter 3.

[6] Usually one does not model the constituents of a word, as the word is taken to be the lowest element in the modelling process. It is assumed here that the meanings associated with a word are *held* in a

defined by the particular search model and not the conceptualisation presented here. A state has associated with it a set of statements whose truth values are not necessarily easily deducible if deducible at all. For example, if the statement 'this search scenario will end in 30 steps' is associated with the general search state, that includes a particular agent whose interactions may not terminate, then whether it takes 30 steps or any finite number of steps is uncertain prior to the scenario. The stack diagram aids in visualising state and substates where if a state represented by a rectangular region influences another then it is drawn adjacent to it.

The visual approach taken is to presume in the style of modular programming, linear computation and network communication that the most general substates are such that they can usually only influence two other states which are drawn above and below it in the diagram as in Fig. 5.3; any other influences the state may be depicted using directed arrows that denote the flows. Flows are static links that indicate the potential of one component to influence another[7] whereas a *path*[8] is a set of flows with a sequential relationship. Visual adjacency between states implies that they will potentially influence one another, or in other words that there are flows between them.

Following sections of this chapter treat the dynamics of search more precisely. It may seem that there is a strong assumption inherent in the stack if one assumes the arrows must all stay *inside* the rectangular shape. The stack representation seems to force an order on the set of flows by putting the components in order and implying that changes/flows in one component must precede changes in the adjacent or prior component. In order to accommodate this it is said that the arrows may be drawn around the components, skipping them, and not always going through them, these cases are illustrated in Fig. 5.3. Even though the apparent forced order may be compatible with many search processes there should not be such a restriction in reasoning about general cases. For example, if the flow in some component is not clear but it is for all other components, then it is convenient to skip detailing that component while addressing and reasoning about others about which information is known.

---

separate entity and do not 'constitute' the word, and thus the sub-state structure here corresponds to that of traditional retrieval modelling.

[7] It may indicate an influence that can potentially occur in a search process without relating to a particular time or a general association that is explicitly time and search process independent.

[8] This refers to ordered paths; there is also an unordered path that is a set of flows without any order.

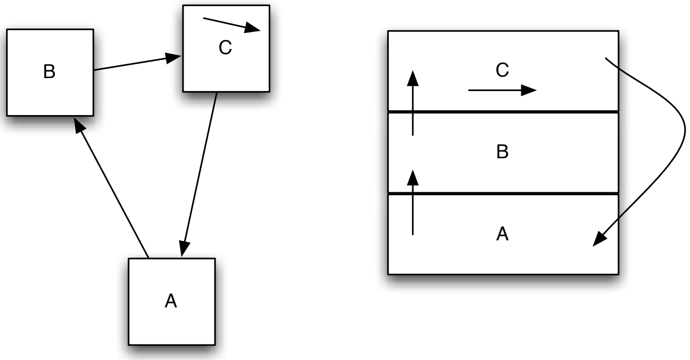


Figure 5.3. Two different visualisations of the same agent, showing components and two types of internal flows. The general representation on the left and stack on the right

The stack diagram in the general case could be any complex 2-D (or higher dimensional) shape where the substate structure could be of a complex type where the influence between substates requires them to be drawn with a non-trivial and non-rectangular topology. We could require that the rectangular regions be adjacent to more than two others, introduce polygons resembling a Voronoi diagram [AOC00], or more generally by also allowing curved (Fig. 5.4) and not-necessarily continuous regions[9]. The point here is that the visualisation techniques suggested in conceptual spaces in [GÖ0] and others are not excluded from our framework as we want to accommodate several methods of modelling cognition; specifically, the 'anomalous state of knowledge' in [BOB82a], 'state of potentiality' in [AG02] and conceptual spaces approaches, so that concepts depicted by those techniques can be represented in our framework without going 'outside our framework'. The state based framework borrowed from QT accommodates the visual techniques for designing software, specifying cognitive functionality, and the diagramming of physical models. As the agents involved are the user, system and researcher, and with simulation requiring the us to implement models (if possible) in software, it was thus necessary to elaborate that the framework and notation to follow are not unfamiliar to the modelling approaches associated to each of these techniques. The following section elaborates the concept of a state showing its structure, notation, with small examples of applying the notation to describe the static aspect of traditional scenarios. The dynamics of a search process are addressed next with the concept of flows, upon which there is a discussion of techniques for approaching the

---

[9] Where the discontinuity could mean that the researcher knows about two different parts of a state and how they influence other states but does not know about other parts.

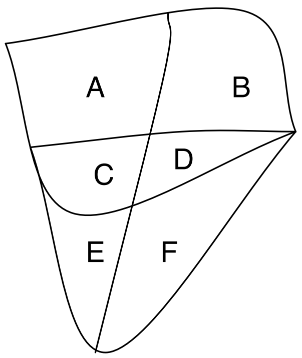


Figure 5.4. A non-linear stack of six components each with three or five adjacent components

multitudes of pre-existing languages partly due to which the definition problem exists. The aim for the proceeding is to go from these simple box drawings of stacks to a refined representation (and notation). The notation introduced is then used to address some applications in Section 6.1, particularly for describing structured document retrieval and user modelling. Finally Section 6.8 argues about the position of the framework in terms of its uses and particularly about the concept of *modelling research*.

5.3 Search Scenario Statics, States and Ontology

5.3.1 Researcher Types

This section follows on from the discussion of states and substates in the ostensive chapter. Recall that a research oracle is an agent that knows all possible future states of a particular search. The optimal researcher R^* is a hypothetical entity that knows all statements and their truth values for any search. R^* denotes an abstract device, a perception of search by which it is meant not only a particular search scenario but any search scenario. If the perception[10] of R^* is applied[11] to a particular search scenario then truth values of all statements pertaining to that scenario would be known. R^* is the largest element in the set of researchers $\mathcal{R}$ that is partially ordered according to the knowledge of an element so if some $R_1 \preceq R_2$, $R_1(\Pi) \preceq R_2(\Pi)$ then it is said that R_2 knows at least as much as R_1 which

---

[10] The words ‘perception’ and ‘framework’ are from here on assumed to be synonymous and the word researcher taken to be an entity that generates or refers to a perception or framework.

[11] Application of a perception means using a particular conceptualisations and models attributed to the perception, such as vector spaces for the geometric perception of a data corpus.

will be expressed by the relation of knowledge potentials $K(R_1) \subseteq K(R_2)$. The *knowledge potential*[12] is the map $K : r \in R \mapsto (s \in ST, [0,1])$ that outputs a set of statements with truth values from the hypothetical set of all possible statements. If applied to a particular search process Π the statement $(R_1(\Pi) \preceq R_2(\Pi)) \longrightarrow (K(R_1(\Pi)) \subseteq K(R_2(\Pi)))$ denotes that the second researcher *has the potential* to know more statements (and their truth values) about the search process than the first over the course of the entire process Π. Note that at some particular time t the relation may not hold so that it is not the case that $(R_1(\Pi) \preceq R_2(\Pi)) \longrightarrow \forall(t)(R_1({}_t\Pi)) \preceq R_2(\Pi))$[13]. Whether this potential is met and R_2 does indeed know more about Π than R_2 upon the search process terminating depends on how the process and researchers method of perceiving it are defined[14], thus the order of researcher is a purely theoretical one and may not be realisable due to practical factors. The meaning of $R_1 \preceq R_2$ is then that the second researcher has the potential to know more than the first for any search process so that $R_1 \preceq R_2 \longrightarrow \forall \Pi \in \Pi^*, R_1(\Pi) \preceq R_2(\Pi)$ but the converse is false as a researcher knowing more in one search process does not imply the same for all processes in Π^*. All search processes Π^* are necessarily known to R^*, which is the end element in the order of researchers such that $\forall R, R \preceq R^*$, the research oracle of the ostensive chapter is one such R as is the external oracle. This includes the case where the search scenario includes humans, thus R^* is clearly not physically realisable as one cannot fully know all concepts in human cognition.

Let R^*_{sim} denote the the optimal researcher for simulated search scenarios so that the potential $K(R^*_{sim}(\Pi))$ holds all knowledge deducible for simulation of Π; the 'research oracle' from Chapter 3 that knows all possible states at some point in a process without knowing which state will be picked next is the weaker perception $R_{res.orac} \prec R^*_{sim}$. As for the self-referential question about articulating who then *knows the knowledge potential* and further elaboration about the set of all statements, these are discussed in Section 5.5, the rest of the chapter takes these theoretical elements as primitives.

---

[12] Knowledge potential will also be referred to as knowledge in this chapter and the context should be used to deduce if it is meant as current knowledge or potential knowledge.

[13] This applies when one can know what can be known by perceptions on the outset, for example the (QT inspired) perception in [vR04] can know all that can be known by the canonical vector space model since the vector space model is included in the Hilbert space.

[14] The tools for perception need to be created to model and express the perception in terms that can be formally analysed. For example association between documents having geometric properties, such as the similarity metric is expressed by a vector space representation.

The researchers R_{vect}, R_{log}, R_{prob}, and R_{KVR-QT} refer to frameworks using the traditional vector space, logical, probabilistic, and QT inspired (as in [vR04]) perceptions of a search process. The abstract device[15] presented by R_{KVR-QT} includes these views so that

$$[R_{vect}, R_{log}, R_{prob}] \preceq R_{KVR-QT} \tag{5.1}$$

Each of these perceptions refers only to the matching and data representation elements, it leaves other elements such as the interface undefined as if they are blind to those elements. In fact, R_{KVR-QT} refers to an independent representation of the search scenario that is deduced upon analysis of the three canonical models of retrieval as well as QT models of physics. The representation was arrived at upon observations of these other models, these observations are search processes, specifically re-search processes. Let the binary relation $R_a \lfloor_{\Pi} R_b$ denote the search where R_a is observing (searching, re-searching) a process Π in which R_b participates. For a researcher to observe a search it must be able to perceive true statements in it, and for it to perceive an agent in that search there must be common knowledge between the observer and the agents being observed thus $R_a \lfloor_{\Pi} R_b \longrightarrow \exists t, K(R_a({}_t\Pi)) \cap K(R_b({}_t\Pi)) \neq \emptyset$ and specifically the knowledge exists prior to the start of the process $K(R_a({}_{t=0}\Pi)) \cap K(R_b({}_{t=0}\Pi)) \neq \emptyset$. For example, in the traditional IR evaluation experiment the researcher knows what agents are to be involved in the experiment. For all other cases it can be stated generally that any observation of a process involves observing behaviours of some agent, thus some knowledge of what an agent is and what to look for is precluded in the observer. Let $R_a \lfloor R_b$ denote the notion[16] that whenever R_b is involved in a search process that R_a has the potential to be able to observe it, thus one can say $R_a \lfloor R_b \longrightarrow K(R_a) \cap K(R_b) \neq \emptyset$. In the perception of the QT inspired generalisation of the canonical retrieval models one can say that due to the Hilbert space of QT *containing* the canonical vector space of IR, any search process represented in the traditional vector space model can be modelled in QT's Hilbert space[17], similarly for the logical and probabilistic models. If R_{trad} is taken to denote the research approach that

---

[15] Abstract device, framework, researcher, and perception are all taken to be similar concepts in this chapter.

[16] Note the usage of the 'observes' symbol without a process subscript to denote observations of all processes.

[17] This is from the purely mathematical point of view, as opposed to conforming to additional criteria of using the Hilbert space according to how QT uses the Hilbert space, as elaborated in Chapter 4.

is the highest in the partially ordered set of all traditional approaches to IR with respect to the matching and data representation components, then whether $R_{trad} \preceq R_{KVR-QT}$ is the crucial research question for [vR04]. It is known that $R_{KVR-QT}\lfloor[R_{log}, R_{prob}, R_{vect}]$, that the canonical perceptions can be understood in terms of the QT inspired perception, but the extent of this is unclear. Let R_{thesis} denote the composite of all the perceptions, models, understandings and theoretical devices presented by this thesis, the rest of this chapter builds onto $R_{thesis} \preceq R^*$ further notation to elaborate on how one can perceive the search process so that $[(R_{lab}$[18]$\preceq R_{trad}), R_{KVR-QT}] \preceq R_{thesis}$[19] which is necessary for a formal framework for analysis of IR concepts from where practical systems can be built.

5.3.2 States, substates and statements

5.3.2.1 The Structure of the General Search State S_Π

For each researcher observing a process Π the symbol ${}_tS_\Pi$ refers to all knowledge the researcher has of the process at time t, thus ${}_tS_\Pi$ is a representation of the statements $K(R({}_t\Pi)) = [(s_i \in \Sigma_\Pi, v \in [0,1])]$ for $i \in 1 \ldots |\Sigma_\Pi|$ where Σ_Π is the set of all statements pertaining to the search process. Given a particular perception R_i one does not immediately know what this researcher can know or not, thus the complete knowledge potential $K(R_i)$ is not fully known, neither is it fully known for some arbitrary search process ${}^{R_i}S_\Pi$, instead one may require to do the investigation specified by $R\lfloor_\Pi R_i$ in order to understand the knowledge potential of perception R_i[20]. A state S_Π can exhibit a representation and the terms of this representation elaborate the perception of the researcher whose knowledge is referred to by that state. The general search state will be denoted S_Π and should be taken to mean ${}^{R_{thesis}}S_\Pi$, i.e. the state S_Π according to the perception or conceptual framework of this thesis. For example, if using R_{KVR-QT} then ${}^{R_{KVR-QT}}S_{t\Pi}$ means R_{KVR-QT} (as it stands), thus it refers only to the weights of documents (according to their relevance) at some point in the search and is written in terms of a convex combination of document

---

[18] The laboratory model as it is called in [IJ05]

[19] Where R_{lab} is the typical IR experimental setup as referred to in [IJ05] that also implies particular software architecture and is referred to here to indicate that our framework must be able to generate practical insights.

[20] The investigation $R_i\lfloor_\Pi R_i$, which pertains to 'self-analysis' (by the researcher), can also be done.

elements (as this is one of the key models in [vR04]). Ideally a state representation is complete, meaning that whatever is observed in search can be associated in some way with the terms of the representation. As it stands, R_{KVR-QT} is not complete, and nor does it aim to be, yet to address the definition problem completeness is desired. For R_{thesis} it is assumed instead that our *state does hold all knowledge about the process but that this must be extracted in some way*, so instead of an incomplete state it is said that there is initially an *incomplete representation of the state*. Let $\mathcal{L} : S \subseteq S_{\Pi} \mapsto L \subset K(S)$ be a function that takes a state S and returns statements about its representation language which are also naturally part of the knowledge of the state. A complete representation would be one from which it would be possible to deduce all possible statements about the search process (at that time) as if one knew the translation map $T : \mathcal{L}({}_tS_{\Pi}) \mapsto K\left(R\left({}_t\Pi\right)\right)$. Complete representations are not expected to exist unless one can categorically know $K(R({}_t\Pi))$ in which case the trivial complete representation would be $K(R({}_t\Pi))$ even though it may not be of much use in that form for the purposes of formal analysis or software implementation. Languages for representation are discussed in Section 5.5. In these new terms it can be said that investigations or researches of search with respect to a particular search process (i.e. that involving a particular system), or research evolution, is equivalent to the process of refining the representation of state and trying to understand $K(R({}_t\Pi))$ and finding the representation which is of optimal use for formal theoretical analysis of the search process (paper-based), analysis by simulation through software or for implementation for a search system itself.

Defining states as entities in which are 'inscribed' all possible research findings within the framework is a notion inspired by QT where the state of a system is represented by a wave function which is a theoretical device (incompletely) representing any and all knowledge about the system to which it refers. In order to obtain this knowledge there require to be techniques of extracting (or querying[21]) which in QT takes place in the Hilbert space representation in terms of projections of spaces onto other spaces. In the vector space model information about relationships between documents can be *extracted* by performing a vector operation between the subject documents; this is since the subjects have representation and a measure (the vector operation) exists in the representation to

---

[21] Asking a question about a physical system amounts to making a measurements which in turn amounts to geometrical operations on the Hilbert space. Asking a query is a 're-search' (see Section 3.3).

assign empirical values to relationships. Several ways to extract relationship information between states and processes are suggested in R_{thesis} (see Section 5.4.4) but these are only guidelines for creating increasingly specific representations of states to produce more useful measures.

In an IR evaluation process with real users, the researcher can learn statements from users that no simulation could predict (in general). However, what knowledge can user simulation bring that is not available through traditional user evaluations? Rephrasing in terms of knowledge potentials, consider the set $K(R_{simulated}) \triangle K(R_{trad})$[22], what statements are in this set? Rephrasing in such terms is an informative way to pose these questions about IR. The usefulness of simulation experiments as referred to by the second question denotes a crucial re-search investigation due to it rising in literature as a significant alternative evaluation route. More generally, one major research goal is to deduce the power and limitations of a particular research approach which begins initially with deducing $\triangle_{(R \in \mathcal{R})} K(R)$ for some known $R \in \mathcal{R}$; this is further addressed in Section 6.8.

5.3.2.2 State structure

The search state is a union[23] of other sub-states where complex relationships between states are addressed by using flows and thus no complex combination functions are required so one can initially ignore the product space technique of combining systems in QT[24]. A flow or 'change' (semantically speaking) is a directed association between two states[25] to denote the influence between them. The given definition is sufficient for the references to flows in this section while a more specific definition is given in Section 5.4.

The general search state of search process Π is ${}_tS_{\Pi} = {}_tS_{system} \cup {}_tS_{user} \cup \ldots$ where ${}_tS_{user}$ and ${}_tS_{system}$ are also complete states (with initially incomplete representations). The substate structure can be of arbitrary complexity in terms of influences (flows) among states which detail all relationships between states, otherwise in terms of states only, it

---

[22] Recall that symmetric difference $\triangle$ of two sets is the set of elements which are in one of the sets, but not in both, i.e. $A \triangle B = (A \cup B) - (A \cap B)$ (see [Ros99]).

[23] As this is a weak association between objects, that of being a part of the same larger entity.

[24] This refers to the use of the tensor product of $\mathcal{H}$ spaces as employed by the fourth postulate of QT [NC00].

[25] A flow can also denote influence between statements by extension since a statement is always corresponds to some state.

is said that, any state is a simple union of substates as opposed to some other complex combination function. This is inspired from QT since sub-states of a state are taken to be in superposition (i.e. in the central modelling method of Chapter 4), everything is always in a 'state of potentiality' (to use the terminology of [AG02]). These states hold all knowledge about a particular entity from where one can extract all that is knowable about an abstract user of type `AU` and an abstract system of type `RA` respectively within this particular search process at time t. Let $S_{\mathbf{T}\Pi}$ denote the incomplete representation of a search process that particularly addresses members of $\forall K(R({}_t\Pi))$ that refer to *time-invariant statements* and *general facts about the search which do not refer to time*. If one omits the reference to time and writes S_{Π}, then this denotes, for the sake of generality, the search state of some process and this could imply either $S_{t\Pi}$ or $S_{\mathbf{T}\Pi}$.

Recalling the generalisation of the ostensive chapter that resolved the conceptual problem by re-defining all processes as search processes, take R_{QT} to mean a QT perception of the physical world and take $R_{QT}(\Pi)$ to mean a framework for interpreting a general process Π. In the QT framework R_{QT} and the QT inspired framework R_{KVR-QT} [vR04], the internal structure of states corresponds in some way to the structure of sub-spaces of a Hilbert space, and of course, it is an incomplete representation and it is not capable of representing all the information from a search process[26]. The sub-state structure is according to traditional retrieval for several common elements that are usually modelled and/or retrieved and it is $S_{\text{word}} \subset S_{\text{doc}} \subset S_{\text{corpus}} \subset S_{\text{system}} \subset S_{\Pi}$. A simple interface structured according to an object oriented implementation language would be represented $S_{\text{button}} \subset S_{\text{window}} \subset S_{\text{interface}} \subset S_{\text{system}} \subset S_{\Pi}$, a user model with 'conceptual spaces' $S_{\text{descriptor}} \subset S_{\text{dimension}} \subset S_{\text{concept}} \subset S_{\text{user}} \subset S_{\Pi}$, and a user model which considers interactive behaviour as $S_{\text{hand}} \subset S_{\text{physical}} \subset S_{\text{interactive behaviour}} \subset S_{\text{user}} \subset S_{\Pi}$. For completeness, in accordance with choosing to associate a state with a researcher, it is said that each substate of a state also corresponds a perception, framework or researcher (equivalently), thus the dimension substate $S_{\text{dimension}} \subset S_{\text{concept}} \subset S_{\text{user}} \subset S$ is said to hold knowledge in the perception or framework of conceptual spaces (as elaborated in [Gö00]).

The definition of S_{Π} as an entity holding all knowledge about the subject (in this case Π) applies to all substates so it is said that all knowledge (knowable by this framework) about

---

[26] Meaning that it cannot completely represent all the information contained in S_{Π}.

a word is contained within the state S_{word} that is a substate of a document, and similarly for other substates. In the traditional sense this means that all the meanings of the word including whatever linguistic analysis can reveal is held by the state[27]. Only R^* knows the complete representation of S_{word}, otherwise for each R that includes the word state in its perception and in S_Π, their incomplete representations are not all of equal usefulness[28]. Whether a word is semantically linked to another is something one can only 'know' about by comparing the representations of the words. Each state can be represented trivially as a set of statements $[s \in \Sigma_\Pi]$, or with a higher level functional grouping of statements $[\text{model}, \text{semantics}, \ldots]$ where $\text{model} = [s \in \Sigma_{\mathsf{model}} \subset \Sigma_\Pi]$. What is the representation of statements? In traditional IR, a corpora is a set of documents so each document is represented by statements like s_a = 'i am a graphic' (which is an implicit pointer to a binary data type[29]), $s_b = [s_{w_i}]$ (a bag of words) or $s_c = [(s_{w_i}, \prec)]$ (a bag of ordered sets corresponding to structured topics) where the statement s_z=['i am a vector space for $[\Sigma_a, \Sigma_b]$']^{30} is the representation in terms of a vector space for the preceding statements.

For each state type there are a set of associated statements, in some arbitrary language, so for the above one can have for state representation language $\mathcal{L}(S_\Pi)$ a QT and technical English for IR, technical English for Computer Science in addition to pseudocode and/or programmatic language. This would also similarly be the case for each substate of the general search state which need not be in the same language. The way different languages fit together for one to define a full state may not be easily specifiable as it is *in the researchers mind or inscribed in the researcher's state*. Let T be a translation $T : \Sigma \longrightarrow \Sigma$ then $\exists T(K(S_T) \subset S_{researcher})$ such that $T_1(R_{QT}) \rightarrow T_2(R_{IR})$ meaning that the researcher has figured out a way $[T_1, T_2]$ (translations) to associate a QT phenomena and search elements but this association is in his/her mind and is necessarily ready to be developed into an operational theory, this is part of the definition problem. In QT, to test whether a statement is true for a physical system[31], a representation for the statement in terms of

---

[27] This refers to word meanings and natural language analysis.

[28] For example, the representations of S_{word} for a language modelling frameworks exhibits more knowledge for a particular process than the three canonical models usually do.

[29] This statement implicitly refers to another statement that elaborates on the content of the graphic, perhaps describing it in binary form.

[30] The English description of a vector space here is taken to mean the complete mathematical description of the vector space for preceding statements.

[31] Which is synonymous with a search process in our generalisation of the search concept.

the representation of the system would be devised then compared through mathematical functions that are geometrically equivalent to projections of one space onto another[32]. In R_{thesis}, this is done through discussion and derivation[33] until there is a way to translate the representations concerned into a mathematical framework where function application would replace the discussion in technical English language, translations are addressed in Section 5.5. In order to relate back to HHIR in academic research, one can denote by S_{Π} the collective knowledge of a research group over the process Π which runs until the group no longer exists and each substate $S \subset S_{\Pi}$ is then the knowledge of a researcher or subgroup of researchers, influence between researchers over Π corresponds to process dynamics as addressed in the following section.

5.4 Dynamics, Flows and Epistemology

5.4.1 Introduction

The last section dealt with the state of a search at some point in time, this section introduces the notation of flows and paths for denoting how the search state changes over time. A flow is a directed association[34] between two not necessarily distinct states, it is a static relationship[35] that indicates which state can potentially influence some other state and is defined as the pair $(S_1 <: S, S_2 <: S)$ where $<:$ is the subtype relationship[36]. Since the information need of an agent is a potential for change in that agent, according to the higher-ostensive principle, a flow also indicates a potential for change of a state, specifically, by influence or interaction from another state (that corresponds to some agent). If two states interact in some way then it is said that the *information need* of these states change. As first explained in the ostensive chapter the external oracle causes an initial change of

---

[32] This geometrical equivalency is in reality only true for a subset of the representations of a system.

[33] As opposed to geometric projections and similar techniques.

[34] The direction, as is explained later, allows the model to distinguish between a change in a system and the doer of the change which would be another agent. Directed associations are therefore implicit in the concepts pertaining to interaction of agents in Chapter 3.

[35] This means that the relationship exists before the system changes; the flows denote potential changes, which by definition must exist before the changes actually happen.

[36] Thus the objects that a flow joins as states and the flow in this example originates in the first element S_1 ending at S_2.

state in the search process that starts the process. This initial change of state is described as an external[37] flow γ_{init}, an interaction from the external oracle that directs into the general search state thereby initiating it. Simulation is the assumed physical setup in all our references to search scenarios unless otherwise stated[38], and in this type of search scenario γ_{init} denotes the researcher clicking the start button in simulation software. This section initially addresses a physical semantics for changes of state or flows borrowed from the concept of energy in physics going on to discuss the equivalence between a flow, energy transfer, information need, information experience, mathematical function application and a programmatic function call.

States that influence only two other unique states are those that can be naturally indicated by a stack model, where for each layer the state for that layer is the combination of all substates of that layer: $S_{L_1} = \bigcup^{L_1} S_i$[39]. If the components of a state are $[{}^{A}S_i, {}^{B}S_i, {}^{C}S_i] \subset S_i \subset S_{\Pi}$ as indicated in Fig. 5.5 and the future state for top component is ${}^{A}{}_{1}S_i \neq {}^{A}{}_{0}S_i$, if it is said that a state change in component A occurred. Usually a change in a substate implies a change in the superstate but it depends on the perception of the observer, if the observer is the user R_{user} (in a search) then component B may not know about change ${}^{A}{}_{0}S_i \longrightarrow {}^{A}{}_{1}S_i$; if agent X (from Fig. 5.5) is a system it only knows the interface ${}^{C}{}_{0}S_i$. If the user is simulated with state ${}^{R_{sim-user}}S_{\Pi}$ then the researcher agent is assumed to know it completely. Thus, sometimes the system can tell that the general state of search has changed but not how it changed, so it can detect changes in statements (or *changed* statements) from the set $[\Sigma_A, \Sigma_B, \Sigma_C]$ without knowing about the identity of the particular changed component.

For any base state (at $t = 0$) each possible future state is caused by a set of pre-future-state changes or flows, different future states are distinguished by differences in pre-future-state flows. An agent may never know an absolute set of all pre-state change flows except in a complete simulation when one should, hypothetically, be able to know all of them (given time).

---

[37] External relative to the system being modelled.

[38] Especially since many of the scenarios discussed are of a hypothetical nature and difficult to conceptualise in the traditional sense but not so in the context of simulation or abstraction.

[39] The union of states here should be interpreted as union of sets of states where each set pertains to a component; each components truth statement is a substate (or equivalently, a subset) as introduced in Section 3.1.

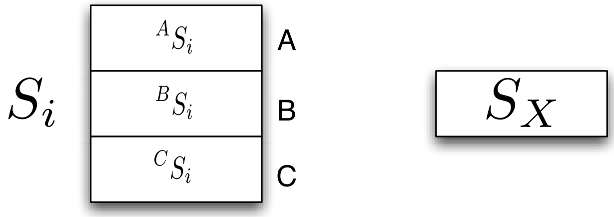


Figure 5.5. Substates corresponding to components of a system displayed as a stack

5.4.2 Semantics

Change Energy is a virtual substance residing on the `AU` and `RA`. Its effect is to cause changes within any agent in which it resides, and thus within any substate of the search state S_Π. This definition of change energy is compatible with that for information as defined in [Mac69] as 'that which causes representations to change'. Although the change energy defined here and information defined in [Mac69] are conceptually equivalent, the notation is kept distinct to show the relation between QT semantics and the IR semantic of information need as discussed in the ostensive chapter. During Π, change energy flows between agents (external flow Γ_E) and within agents (internal flow Γ_I). Each process thus has a set of flows associated with it ${}^{\Pi}\Gamma = {}^{\Pi}\Gamma_E \cup {}^{\Pi}\Gamma_I$[40] and each agent has a set of *interpretation flows* and a set of *expression flows* denoted by $\Gamma_{INT} \subset {}^{\Pi}\Gamma_{Agent}$ and $\Gamma_{EXP} \subset {}^{\Pi}\Gamma_{Agent}$ respectively. Flow of change energy has a directional component which specifies the origin (source) of the change and the last place of effect (sink)[41]. Within an agent or component[42] (where $X :: Y$ indicates that X is an instance of [or is of type] Y) are internal flows to denote changes at finer levels. So it is said that change energy flows between and within agents and components[43]. A flow of change energy is a semantic for the changes incurred by agents at some point in a communication process Π. As S_Π holds all knowledge about a search process including that of changes, for a knowledge potential that also applies to states and any set of objects, $K : S \in S_\Pi \mapsto (s \in ST, [0,1])$, it is said

---

[40] The initial change caused by the external oracle $\gamma_{init} \in {}^{\Pi}\Gamma_E$, even though knowledge of its source is $\notin K(S_\Pi)$.

[41] Directed flows are a way to interpret change except in cases where thinking of energy flow as undirected maybe more suited. The peculiar cases where change cannot be organised as directed flows are not considered here but are seen as flows of an otherwise undefinable nature.

[42] Recall that an agent is represented by some state $(S_{Agent} :: S) <: S_\Pi$ (where $X :: Y$ indicates that X is an instance of [or is of type] Y) and similarly a component ${}^{Agent}S_{comp} :: S$.

[43] That is, the source and sinks can be located within the same agent.

that $K({}^{\Pi}\Gamma) \subset K(S_{\Pi})$ such that $\exists S_{\Gamma} \subset S_{\Pi}$, i.e. that there exists a corresponding state that holds all knowledge about flows for a particular process.

A path is a set of flows, used to describe on a low level any sequence of effects of change energy or *change events*. If ${}^{\Pi}\mathbf{P}$ is the set of all possible paths as 'stored' in ${}^{\Pi}S_{\Gamma}$ in a process Π then in the largest case this is the set of all subsets of ${}^{\Pi}\Gamma$. However it may be the case that some combinations of flows would not occur in any search process so realistically the total number of paths would be far less than $|\wp\left({}^{\Pi}\Gamma\right)|$[44]. Therefore ${}^{\Pi}\mathbf{P} \subseteq \wp\left({}^{\Pi}\Gamma\right)$ paths can be used to describe (on a higher level than flows) the observations that can be made about communication between the agents.

A set of paths are all that is needed to completely describe thus far, any communication process Π. The change events included in a path are not necessarily order independent. It is useful to discuss changes in a process in a general sense without considering orders but when analysing cause and effect the concept of an ordered set of paths becomes necessary. Let an ordered path be a sequence of flows of change energy. Let ${}^{\Pi}\widetilde{\mathbf{P}}$ be the set of all possible ordered paths in a process Π, then in the largest case this is smaller than the set of all subsets of ${}^{\Pi}\Gamma$[45]. However, it may be the case that some sequences of flows are not *valid*[46] so that there maybe a lesser number of reasonable paths than can be inferred from the power set of flows therefore ${}^{\Pi}\widetilde{\mathbf{P}} \subseteq \wp\left({}^{\Pi}\Gamma\right)$ and an ordered path $\widetilde{p} = \left(p \in {}^{\Pi}\mathbf{P}, \prec\right)$ which is the binary order on the set of unordered paths.

A semantic pertaining related to a general path and an ordered path is that for every ordered path there is a corresponding un-ordered path, namely the same path without any ordering. If there is a valid[47] order of flows then to this corresponds an ordered path and a valid unordered path therefore $|{}^{\Pi}\widetilde{\mathbf{P}}| \leq |{}^{\Pi}\mathbf{P}|$ as perhaps not every path can be ordered in valid way, by some definition of valid. An ordered path gives an idea of the sequence of change events in a process, it thereby also introduces the concept of time. The aim is to be able to express and discuss events that occurred in a period of time relative to some

---

[44] Recall that $\wp(X)$ is the power set (set of all subsets) of set X which in this case means the set of all subsets of flows for a process Π.

[45] Although it is unlikely to resemble a realistic scenario.

[46] For example, it may not agree with search semantics with respect to the description of search implied by the sequence.

[47] A path is valid if the events on that path can potentially occur. An ordered path is valid if the events on it can potentially occur in the given order. Whether a set of events can potentially occur together depends on the agents and particulars of the process being modelled.

other period of time, i.e. a set of paths relative to another set of paths. For example, if $p_1 = [a, b, c, d]$ and $p_2 = [x, y, z]$ it could be that x can never occur after *d* unless *d* is preceded by *c*, *b* and *a*, in which case $p_1 \prec_{\widetilde{s}} p_2$ ($\prec_{\widetilde{s}}$ is a set order on ordered subsets) and thus there in an order on the set of ordered paths. There is also an order on the set of unordered paths $\prec_s$ which implies that certain combinations of flows precede some other set of flows regardless of the order of the flows within a path. Measures on the sets of ordered/un-ordered paths can be used to add a quantitative dimension to compare, for example, the *rates of change of groups of events* to indicate notions of *process speed*. An ordered path set ${}^{\Pi}\widetilde{\mathbf{PS}} = \left({}^{\Pi}\mathbf{PS} \subseteq \wp\left({}^{\Pi}\widetilde{\mathbf{P}}\right), \prec_{\widetilde{s}}\right)$ or ${}^{\Pi}\widetilde{\mathbf{PS}} = \left({}^{\Pi}\mathbf{PS} \subseteq \wp\left({}^{\Pi}\mathbf{P}\right), \prec_{s}\right)$ is a partially ordered subset of the power set of ordered or unordered paths for some process Π. The symbol $\prec_s$ denotes an order in which the set elements are sets as well, whereas $\prec_{\widetilde{s}}$ indicates an order where the elements of the ordered set are themselves ordered sets. It may indeed be useful to work with sets of sets of paths and thereby think about rates of change of process speed and other semantics that correspond to measures on such complex sets. Path sets are a sufficient level of grouping of change events (flows) for the purposes of this thesis.

5.4.2.1 Interaction

An interaction between agents can be defined as corresponding to a set of flows containing at least one external flow. Therefore, an interaction is part of a process Π, and corresponds to a path $p \in {}^{\Pi}\mathbf{P}$, it is also true that $\left|{}^{\Pi}\Gamma_E \cap p\right| \geq 1$[48]. The representation of what is traditionally understood as interaction or *user interaction*, is in terms of flows, with the semantics being that of change of an agent by influence, in the form of change energy originating from the other agent. Paths and change energy at this stage give a general syntax and semantics for the concept of interaction. This generic method of characterisation is intended as the most initial language of analysis, from which refinement takes place.

Reversibility A flow $S_1 \xrightarrow{\gamma} S_2$ is *reversible* if there is also an opposite flow $S_2 \xrightarrow{\bar{\gamma}} S_1$. A reversible flow corresponds to an undoable change in a system. In R_{thesis}, the unitary transformations for simulation designs in Section 4.3.7 correspond to flows which

---

[48] Meaning that there is at least one influence/interaction between agents.

are reversible due to the mathematical properties of U, whether a U^{-1} is realistic for a particular search scenario depends on the particular design.

5.4.2.2 The Non-Realism of the `AU`'s

Two agents are distinct if during some process Π in which they participate there exists an external flow $f \in {}^{\Pi}\Gamma_{E}$ which has a source in one agent and a sink in the other. This is a specification of the concept of separation from Section 3.7.3 in terms of paths. There may be instances where it is convenient for research to define an abstract user exactly the way `RA` is defined. Such instances would be realised in our approach by considering two agents whose internal makeup would be the same or similar but they would be separated due to their roles (see next section). This is as opposed to looking at these agents as the same agent and using some peculiar specification of internal flow to show their differences. This choice is made at the moment as it seems to be simpler to separate agents in this way, this may not be the case for some unforeseen cases for which the ideas here will have to be revised. An important point to note at this stage is the idea of an interaction: an interaction is not an event which suddenly occurs in some time but a path (of one or more flows) which exists in the set of all paths, as if pre-set or an 'inherent yet invisible until observed' property of the process. This way of thinking is to enforce the artificiality in the thinking of this entire theorisation as is especially true due to the `AU`, as it refers to an abstract and/or simulated[49] scenario. Now it is required to answer what one means by a retrieval agent and an abstract user. It is made clear below that it is impossible to define the user or system independently, that each pair of agents that interact have common knowledge prior to interaction.

5.4.3 Shared Knowledge

Each agent has exposed properties as any complete definition of an `RA` or `AU` exhibits methods by which an external agent can cause flow from it and to it (external), and/or

---

[49] Meaning that the scenario is not necessarily physically realisable but an abstraction of it can be simulated.

within it (internal)[50]. Thus one agent is perceptible to some extent to another agent and they can cause flows within each other. The meaning of perceptible in this context is that one agent knows of the presence of another agent if they are are engaged in a process. Two agents cannot communicate unless they *share knowledge about some particulars of the communication*. First, they should know that they are in a state of communication and are in some kind of process; secondly, that there is at least another agent in the process; and finally that communication means flow of change energy. It is proposed that these three pieces of knowledge are necessary for any communication and only sufficient for simple processes, with the last condition indicating (by definition of flow) that the agents will know that they have the potential to cause changes in each other. Recall that the communication process is being perceived as the researcher, and everything is known about the `AU` and `RA` since they are being designed from foundations, and also, the agents will *know* limited amounts about each other. Therefore, inherent in the nature of the agents is an understanding of their purpose for existence on a general level: to communicate by engaging in a process of influencing the other agent(s); and at least a simple belief about the standard form (or schema) of any other agent[51].

In the path set for any process Π, there will be at least one external flow. Recall that every process therefore exhibits at least one interaction[52]. The interaction in this case is a more general type, as indicated in above discussion. The *weak law of dependence* is that there are no *complete* definitions of `RA` which are independent of references to some `AU` definition. Since the properties exposed by any *complete definition* of an `RA` indicate how an `AU` could affect that property, each method of altering an `RA` property is at least a partial `AU` definition, namely that of an `AU` which at the least simply attempts to influence change of that property. Thus any simulated user model will necessarily imply a corresponding system interface that corresponds to the user model's interaction capabilities and any system interface similarly implies a user model that corresponds to the interactions permissible on the interface. The *strong law of dependence* is that there are no *complete*

---

[50] The agent causing flows directly within another agent is a simulator which has knowledge of the internal structure and is required to pick states at the internal level, as elaborated towards the end of this chapter.

[51] Defining agents and processes in this way is inspired by HHIR, see Section 2.4.

[52] Note this interaction is not be confused with the meaning of interaction as understood in traditional retrieval research.

definitions of a `AU` which are independent of references to some `RA` definition. Since the properties exposed by any *complete* definition of an `AU` indicate how an `RA` could affect that property each method of altering an `AU` property is at least a partial `RA` definition, namely that of an `RA` which at the least simply attempts to influence change of that property. The dependence laws do not prohibit incremental design of agents which in some point in its design have no alterable properties defined. Instead it means that a definition cannot be complete with regards to fulfilling the purpose of communication, until it specifies how its own properties can be changed. The strong law of dependence is thus named since requiring user models to be tied to interfaces is further from realism (and therefore a *stronger non-realism*) in the traditional sense than the dependence of interfaces to user models.

5.4.4 Measures on Sets

At this current level of abstraction of a search process which still requires refinement to be fully effective in addressing practical search phenomena, there are some general techniques that can be used to analyse changes within search which have been characterised as flows, and higher structures like paths. The semantic of the flow and the path correspond to their meaning in natural-language, at the least it is a more (mathematically) precise way of talking about the same thing. However, it is likely that ideas from the mathematical theory of information and other disciplines can be used to reason[53], at least on a structural level, about these representations. For example, given a set of flows for a process, according to the probability that they will occur, the information content of a path denoting a set of events is $\mathcal{I}\left({}^{\Pi}P_1\right)$[54]$= -\Sigma_{f_i \in P}\log(\mathbf{Pr}(f_i|S_\Pi))$[55] that would indicate peculiarity of the path and functions $f(\mathcal{I}({}^{\Pi}P_1), \mathcal{I}({}^{\Pi}P_2))$ would be used to compare between paths. As the probability is formulated by considering the state of the process, an ordered path or path set can also be hypothetically assigned useful values for the purposes of quantitatively comparing

---

[53] Meaning that the mathematical theory of information could perhaps be applied to distinguish between representations.

[54] Note the overloading of the function $\mathcal{I}$ which was used in Chapter 4 to refer to the dual space of a bra vector

[55] where $\mathbf{Pr}(f_i|S_\Pi)$ would require to be deduced according to some $R \prec R_{thesis}$ for a process Π.

and classifying scenario designs[56]. Thus, the approach is quite simple in that it records changes, however vaguely they can otherwise be defined. One important philosophical point to note is the assumption that inherent in agents is an idea of their purpose and a perception of the basic properties of other agents (purpose for HHIR agents was discussed in Section 2.4). This is the minimum default shared knowledge between the entities[57].

Measures of the complexity of a path corresponding to the semantics of the level of detail/richness of a qualitative description can be formulated, by initially observing the number of flows in a path. However, as flows are of two types, one needs to consider the relative similarities between an internal flow and an external flow in making judgements about the power of the description. If the source of a flow is seen as a cause with the sink being the effect then it is possible that a system of logic can be employed to formally argue about changes in a Π, for example, the flow $\gamma = (S_1 :: S, S_2 :: S)$ whose semantic is that S_1 influences S_2 can be taken to mean the logical implication $S_1 \rightarrow S_2$ which supports the semantic. There are several such routes of analysis which can be employed including the combination of information theoretic measures and of logic[58] to give quantitative measures of logical uncertainty[59]. The meaning or usefulness of any such analysis depends on further specification of the presented concepts.

The main idea is to use this rather basic approach to define in more detail the relationships between flows/changes within agents, thereby incrementally attaching deeper meanings (and subsequent representation) to each change. If a change can be described in more detail such as this example from traditional IR: "the system's interface changed to reflect user interaction" then it is required to deduce how to introduce concepts like 'interface change'. In any process, as the agents are composed of components, the flows will require to be subdivided into other types to distinguish between the changes that happen in one component with the changes attributed to another component and any relationships

---

[56] There are clearly a multitude of techniques for enumeration of flows, paths and ordered paths, assignment of values to them and subsequent measures on these enumerations using the values.

[57] It is unclear if these default characteristics are representable within the same language or if there is a meta-language to specify them.

[58] Not making any assumptions about the type of this logic except it is not restricted to classical logic due to the definition of all physical processes, including QT ones, as search processes thus implying that they should be definable in our framework and any logic of flows describing them would not be classical in the way the logic on the QT Hilbert space is non-classical.

[59] In this regard one could have logical uncertainty principles for each S_{Π} to characterise the flows within that could be useful for user modelling to denote different modes of behaviour.

between them. So far the only relationships between flows are that of belonging to the same path or ordered path, but it is likely that further formalisation can take place in terms of flow types which may aid in theoretical comparisons between agent designs. Given the structure and semantics of flows and paths, there are many avenues for further theoretical development, one set of avenues point towards domain theory [DP02, GHK+03] and similar formal theories for theorem proving while another set of avenues point towards measurements and information theory[60].

5.4.5 Flow Types

Without loss of generality let the internal flows in AU and RA be denoted ${}^{\Pi}\Gamma_{\mathtt{AU}}$ and ${}^{\Pi}\Gamma_{\mathtt{RA}}$ respectively then for Π, ${}^{\Pi}\mathbf{F}_I = {}^{\Pi}\Gamma_{\mathtt{AU}} \cup {}^{\Pi}\Gamma_{\mathtt{RA}}$. If an agent contains two components labelled *a* and *b* then let it be indicated by the expression ${}^{\Pi}\Gamma_{\mathtt{AU}} = {}^{\Pi}\Gamma_{\mathtt{AU},(a,a)} \cup {}^{\Pi}\Gamma_{\mathtt{AU},(b,b)} \cup {}^{\Pi}\Gamma_{\mathtt{AU},(a,b)} \cup {}^{\Pi}\Gamma_{\mathtt{AU},(b,a)}$. In the preceding, flows were limited to be within an agent or between agents, it is now said that if an agent is a composite, the internal flows can be between its components: $f \in [..,..]_{(a,b)} \cup [..,..]_{(b,a)} \subset {}^{\Pi}\Gamma_{\mathtt{AU}}$ and within components: $f \in {}^{\Pi}\Gamma_{\mathtt{AU},(a,a)} = [..,..]_{(a,a)}$ where $[..,..]_{(a,b)}$ and $[..,..]_{(a,a)}$ denote a set of flows from the source to the sink, as specified. Each partition of the internal flows set corresponds to flows between one component and another or within a component, therefore the number of possible partitions is the square of number of components and each partition corresponds to a flow type as summarised below. In general, one identifies every part of an agent that is to be reasoned with and decides if the overall functionality of one of these agents is separate in some way to another, if it is then they are separate components otherwise they are the same. Further, there are no unambiguous rules in general but heuristics, for example, interface and data layers can be two different components or part of one, depending on the purpose of the analysis. Whatever the design of an agent, the benefit of casting it in this form is that it makes it clear and allows structural analysis to be done.

In the above representation, ${}^{\Pi}\Gamma_{\mathtt{AU},(a,a)} = [..,..]_{(a,a)}$ can be expressed as a set expansion of the form $[\gamma_1, \gamma_2, \gamma_3, .., \gamma_k]$ in which each γ_i, apart from differing in subscript, have

---

[60] One of the aims in formalising using the notions of types and such is to inspire a view of the communication process as a series of computations. Thereby presenting a door to formal theorists to apply their tools to the analysis of IR, and specifically to the dynamics of IR.

equivalent meaning at this point. This ambiguity is useful when little is known or assumed about changes in an agent. If more details are known about a flow then it requires to correspond to further specification, which will be explained in the next section.

External and Internal flows are the two most general type of flows represented by ${}^{\Pi}\mathcal{FT} = {}^{\Pi}\mathcal{FT}_I \cup {}^{\Pi}\mathcal{FT}_E$ where ${}^{\Pi}\mathcal{FT}$ is the set of all flow types within all agents for Π, following this the two types of internal flows ${}^{\Pi}\mathcal{FT}_I = {}^{\Pi}\mathcal{FT}_{EXP} \cup {}^{\Pi}\mathcal{FT}_{INT}$. Then to each component expressed as ${}^{\mathtt{AU}}S_i$ for `AU` corresponds a flow type $\mathcal{FT}_{\mathtt{AU},(i,i)} \in {}^{\Pi}\mathcal{FT}_I$. Finally, for any two different component combinations there are flow types $\mathcal{FT}_{\mathtt{AU},(i,j)} \in {}^{\Pi}\mathcal{FT}_I$ and $\mathcal{FT}_{\mathtt{AU},(j,i)} \in {}^{\Pi}\mathcal{FT}_I$. Hence there are $|{}^{\mathtt{AU}}S_i|^2$[61] flow types for agent `AU`; moreover ${}^{\Pi}\mathcal{FT}$ has size $\sum_{k=1..|Agents|} \left|{}^{\Pi}S_{\mathtt{Agent_k}}\right|^2$. This size is valid until each flow type is broken down again into further sub-types as each internal flow type is not necessarily atomic.

Π :: Flow A search process can itself be interpreted as a flow between two states which the observer of the search can have. If the observation is $R\lfloor\Pi$ then analysis of the researcher corresponds to $R_{obs}\lfloor R\lfloor\Pi$ such that R_{obs} interprets R as $S_1, S_2 \in {}^{R_{obs}}K({}^{R}K(\Pi)), S_1 \xrightarrow{\Pi} S_2$. The states S_1 and S_2 simply correspond to the knowledge of a researcher prior to and upon the completion of search process Π respectively.

The semantic of the flow is a change regardless of its type and it can be made more specific by drawing a correspondence between each flow type and the energy that is transferred in the flow. On the highest level, each agent contains *change energy*. The semantic of 'change' in change energy is the most general semantic description of the purpose and function of the energy. When this energy is transferred by means of an external flow it changes type and its source type differs from its destination type. The energy changes type again as it flows between components and also within components, see Section 5.4.6. Although the flow is seen as a syntax and its meaning as the semantic, this is not certain and is only an initial setting. The idea is to refine the syntax and also refine the semantics. For example, using traditional IR ideas in a crude way, a flow can be refined to a representation of a vector rotation, with an IR semantic of relevance feedback. This semantics can be balanced with the notion of energy by interpreting different states of relevance as different types of energy and a change of relevance as a change of energy type; relationship between

---

[61] The set size is only of the component level and not also of substates.

syntax and semantics are depicted in Fig. 5.6.

5.4.6 Structure of Semantics

The most general energy type of a process is a super-type of the energy types of the agents engaged in the process and this is represented by ${}^{\Pi}\mathcal{ET} = \cup_{k=1..|Agents|}{}^{\Pi}\mathcal{ET}_{Agent_k}$. Further, an agent energy type is a super-type of the component energy types of the components within an agent, represented as: ${}^{\Pi}\mathcal{ET}_{Agent_k} = \cup_{i=1..|S_{Agent_k}|}{}^{\Pi}\mathcal{ET}_{Agent_k,c_i}$. Finally, each component's energy type may itself be a super-type such that it's subtypes correspond to sub-components, and this is the structure of semantics. This simple typed view of energies is also a structured view of the semantics of changes. A clarification to note at this stage is that the concept of syntax and semantics are relative and are structured in their own respect. Energy is a semantics for changes in the communication process, but has itself further semantics and syntax. For example, the use of types for energy is a syntax for representing energy. A semantics for it would be a further refinement of the meaning of energy which could be specified in natural or formal language. A formal specification of the semantics of energy could then itself exhibit further syntax and semantics.

Energy conversion is described in two ways, the most general is the structural way, it is specified by a type conversion, and is a syntactical description of energy conversion. The second way is to attach more meaning for each type of energy and show how that description changes in a conversion, this is the semantics orientated description. This refining of syntax, semantics and the drawing of relations between them is a functional view of the research process[62]. A depiction of this process is in Fig. 5.6 which is a map showing where our discussion stands so far and where it needs to go as is indicated by the relevance feedback semantic.

What is the difference between flow type and energy type? Flow types are for structural classification whereas energy types denotes semantics, so for the flow $\gamma_1 \in \mathcal{FT}_1 <: \mathcal{FT}$, $H : \mathcal{FT}_1 \longrightarrow \mathcal{ET}$ denotes all flows of a certain (sub) type that have the same meaning. Flow types and simple semantics could apply to words like 'internal', 'external', 'system decision','user (cognitive) decision' and so on. Therefore *the syntax of changes are sets*

---

[62] R_{thesis} encourages a particular type of research and therefore a particular type of search process.

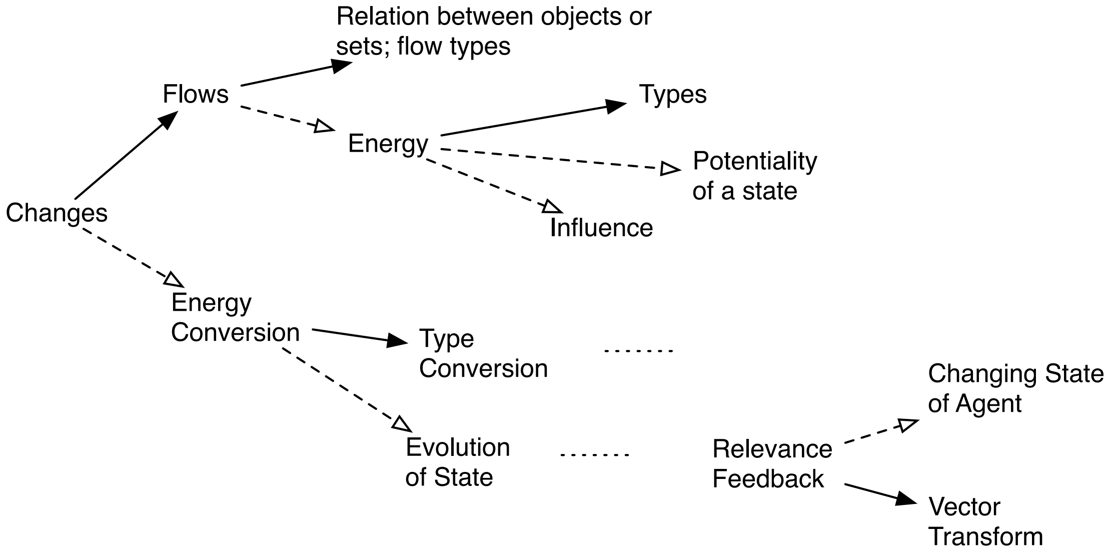


Figure 5.6. A diagram of the research & refinement process showing the initial concepts and the goal concepts, which are yet to be tackled

of flows whose semantics are the denoted by flow type labels. In addition, it is said that *the semantics of changes are energy types* which in turn act as a syntax for semantical meanings. For example, the semantics denoted by ‘relevance judgement has been given’ can be associated with an energy type; this is discussed further in Section 5.5.

5.4.6.1 Constant properties for semantics

Each component contains energy of type ${}^{\Pi}\mathcal{ET}_{Agent_k,c_i}$. A flow is a movement of change energy during which there is a energy of type ${}^{\Pi}\mathcal{ET}_{Agent_k,c_i}$ which is moved to a component where it changes its type to ${}^{\Pi}\mathcal{ET}_{Agent_k,c_i}$ and is part of the combined energy of that component. A flow can be internal to a component in which case the energy type need not change.

Consider some process Π in which an energy type $e \in {}^{\Pi}\mathcal{ET}, e <: {}^{\Pi}\mathcal{ET}$[63] is either never converted throughout the life of the process (in any path) or the proportion of it converted is equal to that which is converted to it (in any path). Thus this energy does not decrease

---

[63] Note that the energy type here also corresponds to a set which without loss of generality is also referred to with the same symbol, as are its subtypes. Thus, the structure of the associated set (in general) follows the structure of the types in the case that the element corresponding to the subtype also belongs in the set of the supertype which would indicate that the semantic corresponding to the subtype is feature of the process rather than a hypothetical relationship. The correspondence between a set and a type used here is used without further elaboration for the flow supertype (and its subtypes).

or increase in any path which is part of the process Π. In such a case, it is said that the energy type $e \in {}^{\Pi}\mathcal{ET}$ is conserved relative to the process. Similarly consider a path $p \in {}^{\Pi}\mathbf{P} \cup {}^{\Pi}\widetilde{\mathbf{P}}$ in which energy type $e \in {}^{\Pi}\mathcal{ET}, e <: {}^{\Pi}\mathcal{ET}$ is either never converted throughout the life of the process or the proportion of it converted is equal to that which is converted to it. This energy does not decrease or increase in any path which is part of the process Π. In such a case it is said that the energy type $e \in {}^{\Pi}\mathcal{ET}$ is conserved relative to the path p. This definition highlights a special case property of energy in which energy is *conserved*. The semantics of energy conservation are very generally stated here without any elaboration of a measure for energy. The conservation principle as a semantic is made use of in the proceeding in discussing the retrieval concept of *information need*, it is proposed that the *need is preserved* throughout a search process. This energy conservation is the *highest level semantic pertaining to a search process* and is in agreement with the principle of higher ostension.

5.4.6.2 State and Energy

In the prior discussion, each component and therefore each agent has a corresponding energy. Energy is a property of a component which is defined to be inherent to all components and agents so that there exists a way, albeit a general way, to assign meaning to changes. The energy is therefore a partial representation of state, such that all knowledge about energy (and the semantics that accompany it) is held by $S_E \subset^* S_\Pi$ where $a \subset^* b$ means that a is either a subset of b or in a subset whose eventual superset is b[64]. A state description contains both syntax and the corresponding semantics.

The state, like flows and energy has a type which can be a supertype which implies that there can be a general representation of a state with the most general type being S_Π; note that the symbol for a state is also being used to denote the type of that state. Types are again used to put a structure to the different ways of expressing a state description. The type structure of states is a copy of the set structure of states with additions, so it is said

---

[64] The notions of change, energy and state have similarities to notions with the same name in physics. These notions are used to express particular components in [vR04] where QT mathematics are used to represent retrieval models and some QT semantics relative to IR are also discussed. In using these notions to talk about communication in general, the enquiry about if QT syntax and semantics are suited to completely model IR is automatically addressed.

that S_{Agent_i} in addition to being state, denotes a super-type of the component state types ${}^{\Pi}S_{Agent_k} = \cup_{i=1..|S_{Agent_k}|}{}^{\Pi}S_{i,Agent_k}$ where ${}^{Agent_k}S$ is the set of component states (stacks) in the k-th agent. The additions referred to here are semantic relationships of types, a set of substates of an agent could be of the same type as a set of substates of another agent, an example would be two IR systems in a search scenario using the same corpus so ${}^{\Pi}S_{c3,Agent_2,} <: {}^{\Pi}S_{c2,Agent_4}$[65]. The state supertype $\mathcal{ST} :> S_{Agent_i}$ contains the agent states types along with any additional types thus part of its structure directly resembles the substate structure of S_{Π}. A state changes over the course of a process but its type will not necessarily change. Energy conversion always corresponds to a state change and state change always corresponds to energy conversion of some type, this means that there is always a semantic for any state change albeit a general one.

5.4.7 Using Flows in a Scenario

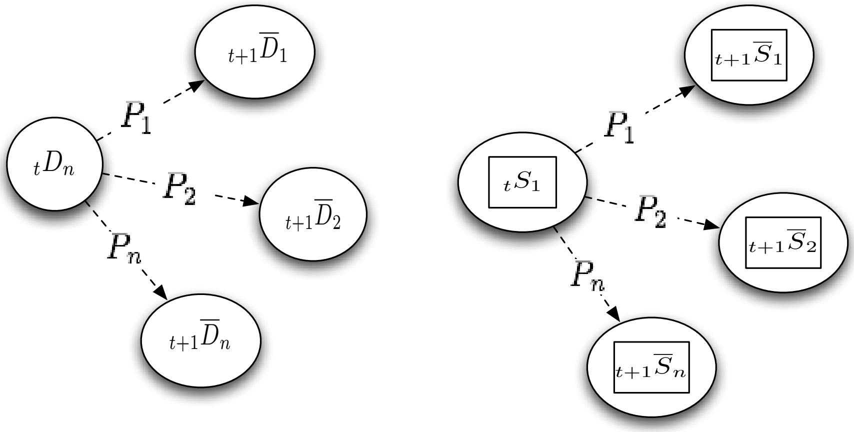
5.4.7.1 Picking States

At the start of a search process *the external oracle is the first mover* that starts the process Π, at each following point in time the next search state[66] is picked from a set of possible future states of search. A state changes if any of its substates change, thus $\forall\hat{S}(\hat{S} \subset^* S)({}_t\hat{S} \longrightarrow {}_{t+1}\hat{S}) \longrightarrow ({}_tS \longrightarrow {}_{t+1}S)$. As the state picking is also hierarchical, at each level of the hierarchy it is said that a flow between substates causes this change to happen thus $Y \xrightarrow{\gamma} X$ denotes that agent or component Y causes change in agent X as a flow between them takes place. The flow that takes place at a sub-level causes state change on that level as well as on all higher levels of the hierarchy. For a visual change to take place like that in a traditional IR interface a set of γ's would be required (unless it's a very basic model) to represent visible state change. Visual state change with respect to user modelling would be if an icon in the system's interface changing from blue to red caused 'some changes' in the user's cognition but it was really the changing from blue back to green that caused the user to interact in a particular way and cause visual state change. In general sense, there is the state change diagram in Fig. 5.7 where the potential

---

[65] Enumeration of components is not dealt with in this thesis, however for structured simulation studies with multiples types of components it could be useful to create a way to consistently number scenarios.

[66] This refers to the most general state.



Ostensive Retrieval Interface State Diagram for an agent stack

Figure 5.7. Visual similarities between ostensive retrieval interface and general state change

next states $_{t+1}\overline{S_i}$ are each caused by a path P_i but in the traditional ostensive model of [Cam00] the visual state change is denoted by a set of documents where the interface indicates indirectly that the documents presented $_{t+1}\overline{D}_i$ are at the top of the relevance list as indicated by the first diagram in Fig. 5.7 where P_i denotes (without loss of generality) the suggested documents for the next step.

As state picking is done by flows where the next states are given if it is said that the current, most general system state is S_{system} (also referred to as $_0S_{system}$), then potential future system states can be denoted $_tS_{system}$. It could be that in each of these future states or a subset of them there is but one difference, such that if the state is decomposed into substates $_tS_{system_i} = {}_tS_{system_1} \cdots {}_tS_{system_k}$ then it is some variation in a state $_tS_{system_3}$[67] that has caused future states while all others are the same such that if this substate were not present in the system then the current system state does not change except trivially (due to time parameter incrementing). An example of a high-level method for state picking is to use the utility measure $\nu(\psi, {}_{t+1}\overline{\psi})$, as introduced in Section 4.3.4[68], over the current state and potential future states to identify the highest scoring future state.

The stack diagram is a static visualisation of the states of a search, whereas the state graph in Fig. 5.7 is the dynamic view. The dynamic view of potential future states is represented by the sum of their possibilities, or in QT terms by a superposition, and a flow

---

[67] Some ordering is assumed on the set of substates of potential future states so that the third substate always refers always to the same *state type*.

[68] Where the ψ corresponds to a state S and $\overline{\psi}$ to possible future states.

then picks a state implying the superposition of potentials then collapses into the picked state, thereby creating a new superposition of future states stemming from the newly chosen state. In the perspective of the researcher employing a random method for picking future states in a user model, this superposition is the anomalous state of knowledge from [BOB82a]. It also corresponds to the potentiality of concepts in [AG02]. This is the way the concept of superposition from QT is used, and it can be expressed in our notation by overloading the semantics of $\cup$ to be a union of states, i.e. ${}_{t+1}S_{\textit{future states}} = {}_{t+1}S_1 \cup S_2 \cup \ldots$. Therefore all unions of states can be interpreted conceptually, depending on the context, as a superposition of states[69].

A further relationship concept from QT could also be used here; the key idea of a (quantum) entanglement between two or more objects is that these objects have to be described with reference to each other, even though they may be different (i.e. in physically different locations), see [NC00, Gri02]. The use of entanglement would apply when states or substates change together or there is always some consistent influence in one from the other as one changes state. One could say with this simplistic definition that *the interface state is entangled with the user's expression state as user's physical actions on the system is always captured*. However, a more conservative definition preserving some of the QT intended meaning would be that *two states are entangled when they not only influence each other consistently but that there is missing knowledge as to why this happens, and/or that they influence each other in a way not fully determined, determinable or deterministic*. In terms of knowledge potentials for two entangled states S_1 and S_2 there is a flow γ, that is $S_1 \xrightarrow{\gamma} S_2$ but current knowledge about this as denoted by the potential $K(S_\gamma)$ is minimal, if one deduces $(K(S_\gamma) \cap K(S_1) \cap K(S_2))$ then the states are no longer entangled but instead *correlated according to the new found knowledge*.

An example would be where a user model generates a user's (simulated) interactive behaviour on (simulating the) observation of certain documents but the researcher does not

---

[69] This has a wide range of implications which are not addressed here. However, note that this overloading is seen as a nature one. In a search system which consists of sub-components the $\cup$ can used to express the system in terms of its components and it can also be interpreted as a *superposition* of sub-components. The latter could be taken to indicate that in a particular future state the components of the search would change ('collapse') into another set of components. It is an open problem as to the breadth of applications of this overloading for IR, one possible example is the modelling of highly-adaptive systems which change their components, such as their query-document matching method or indexing method, according to user behaviour.

yet know the topic common to the documents interacted with thus due to this lack of knowledge the researcher could view these documents as entangled until their precise correlation (i.e. by formulating a common topic) can be determined. Thus the entanglement concept can be employed in this simplistic sense as a modelling tool to represent a lack of knowledge about the precise correlation between documents (or other items).

Relative to the system view an oracle picks a flow to choose a state, at the level of the interface stack the oracle is the user. On other levels of the stack, if the scenario is simulated, the simulation software picks states within the system and user models, and in general for each level the picking is done in one of two ways:

1. Automatically: When deterministic, for example, when there is only one way for the change to go, as with most IR systems
2. Randomly or with a random model which would apply more to a user model to simulate realistic behaviours

In case one above there is some function call which would inevitably have parameter passing, let a parameter passing flow of subtype $\gamma_p <: \gamma$ be depicted by Fig. 5.8 as a flow with explicit parameters. Function application is thus represented as one or more flows that change a set of 'value sub-states'. All flows have parameters that are implicit unless made explicit and are represented by a state associated to the parameter of the flow such that $\gamma_{\omega <: S}$, within ω is embedded all the details of the flow[70].

In case two above it is also an implicit parameter. In fact, even the general flow type has an implicit parameter, as there needs to be something to indicate details about how or what is going to be changed, in this case it is a parameter with the state name (or number) according to an enumeration scheme such as that employed by program compilers. If for modelling purposes the details of flows are required then their parameters would require to be explicitly specified.

Flow Parameter as a State Each parameter in a flow is also a **state** as it contains knowledge, so one could say that it is an agent in itself. This thesis does not consider

---

[70] Implicit parameters are those about which one has knowledge except they contain their sources, i.e. $K(\omega)$ includes statements about the origin of the flow in which ω is a parameter.

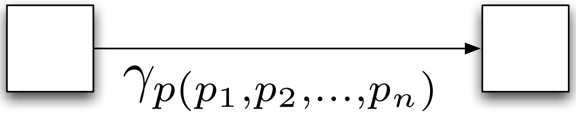


Figure 5.8. Flow with explicit parameters

the concept of flows internal to the parameter state, which would be a flow within the parameter of a flow, except for the most general flow. The most general flow is that of the search process itself Π which has parameter S_Π that is passed between *two states of the observer of search*. A state of the parameter of a flow is said to have the semantic of a sub-energy, this keeps the semantic relationship between a flow (and its energy) and its parameter consistent, so that the energy of a flow is a some combination of the energies of its parameters.

5.4.7.2 Relation to Higher Ostension

State changes caused by flows in the higher levels of the stack are more subtly defined, for example, in R_{trad} it can be said that the flow γ_x denotes the output of a matching function that causes the 'corpus to **experience** a call (a programmatic function call)'. Equivalently change energy flows from the matching component to the corpus and this flow is the interpretation type thus $\gamma_x <: \mathcal{FT}_{INT}$[71]. Also applying to this flow and denoting the traditional function call that matches query with corpus is the concept of higher-ostension as it is said that *the information need of the corpus has changed* and that information has flowed to the corpus from the matching component; thus all these different ways of describing matching are made synonymous. The advantage of several description with flows/stacks being a middle language is that it makes it easier to map to QT (or other frameworks) for formal analysis and to a programming language for implementation. A further advantage of several description of flow and especially that of introducing types is that since a flow denotes a potential event, one could characterise search processes according to flows, for example, one can define a Π as one in which ${}^2\gamma_x, {}^3\gamma_x$ occur and ${}^4\gamma_x$ does not, or more significantly that flows of that *type* happen (or not) and use this as a point of analysis to discuss the process with the aid of the various semantics. The

---

[71] The process superscript is omitted without loss of generality.

semantic of the 'movement of change energy' associated to a flow can be translated as 'function call' or 'programmatic link' or "function pointing' if CS-semantics are adopted (i.e. if translate to CS takes place). An example of a flow type is ${}^{i}\gamma_{x}$ denoting the 'match word to some document..' function, or γ_{m} denoting a function that changes statements associated with the corpus therefore modifying $K(R_{corpus}(\Pi))$ where R_{corpus} denotes the perception naturally associated with the corpus substate of the search state $S_{corpus} \subset^{*} S_{\Pi}$.

In summary, each ${}^{i}\gamma_{x}$ denotes at once, a potential interaction synonymous with influence, change energy flow/movements, information flow and function call. Further, the flow will not be an active interaction except in a process where there is a *path of events* that includes it. This way of specifying the place of flows implies deterministic behaviour. However, it is not restricted to deterministic change - as non-deterministic change is a key ingredient to model the lack of knowledge researchers/perceptions posess especially when the user is modelled where 'changes happen' but the internal reasoning for it are at the least uncertain if not unknown, this is addressed by example in Section 6.1.

5.4.8 Convergence

In the above sections low-level conceptualisation of state changes were developed through specification of flows and paths. In order to understand higher-level changes in a search the concept of convergence is introduced in this section to classify particular semantic stages in a search. The low-level flows and paths are what can be called the 'quantum changes' whereas convergence is akin to 'classical changes' both in the sense of ad-hoc IR and classical physics. Convergence is a concept that is used to model a measuring device. As in QT, a 'measuring device' is an agent that has state as well as observing other agents and their states, thus it corresponds to a researcher R_{eval}. An example of a measuring device from traditional retrieval is the human evaluator who is also an agent. In order to observe other agents the evaluation agent must share knowledge about other agents prior to observing them. In the way an user knows the system interface, the agent or component (or stack) which causes change to another stack also knows something about it.

Convergence or state convergence is when one state knows a certain amount about another in a different agent or stack, through knowledge having been shared over time, either dynamically or statically. If the convergence was static it is said that the states *'converged on start'*.

In the non-deterministic selection of future states that is required for simulation of user models for example, a random function in[72] the simulator picks the next states. A random function is inscribed in a substate of the agent that picks and needs to know some information about what choices there are to randomly pick even if this is simple information about the number of items. For a state in a stack (or the general state of the stack or the general search state), the possible future states can be picked by a random function which is represented itself as an implicit agent whose details are inscribed (along with details of randomness) in the state from which a flow originates to pick[73]. The way it calculates what to pick is denoted by S_r which is the internal state of a stack layer from which the flow for picking originates. The statements of a state a researcher may know could be substates themselves, an example from [Cam00] is a list of documents for ostensive interaction, and thus is the way to classify user models for that type of system/interface.

Convergence is when for agents `AU` and `RA`, $\exists S_1 \subset S_{\mathtt{AU}}$ and $\exists S_2 \subset S_{\mathtt{RA}}$ such that $\exists C = C : (S, S) \mapsto S$[74] a function embedded in an implicit agent outside of `RA` and `AU`, which could be inside researcher or more practically inside an evaluator representing evaluation metrics in a simulation. Such an evaluator would check the value of this function in order to monitor the progress of the search. In the language of QT if $\mathcal{L}(S_1) \subset \mathcal{L}R_{QT}$ and $\mathcal{L}(S_2) \subset \mathcal{L}(R_{QT})$ then convergence (sub-type of the convergence defined above) could be a map between spaces $(d : (\mathcal{H}_1, \mathcal{H}_2) \mapsto \mathcal{H}) <: C$. Thus according to R_{thesis}, any measure between two states regardless of the representation of state, in a language whose corresponding perception precedes $\mathcal{L}\left(R_{thesis}\right)$, is necessarily a subtype of the convergence function. The knowledge potential is the subject that is changed or converges from one potential to another, thus it is the parameter of the convergence function $X <: K$ such that $C : X \mapsto X$. Also there are other operations that are in fact of the same type as a convergence function as listed below, notice that therefore several seemingly different things can be **checked** by functions of supertype convergence, and therefore it is said that

---

[72] The random function is in the knowledge of the evaluator.

[73] Usually the simulator simulating the scenario, i.e. R_{sim}.

[74] The convergence function takes two states and returns a state whose knowledge includes details of relationships between the input states.

they are problems pertaining to convergence:

1. System understanding problem (SUP), denoted $C : (S_{Agent}, S_{Agent}) \mapsto S$
2. Simulation knowledge (the gain and change of it), denoted $C : ({}^{R_{thesis}}S_{sim}, {}^{R}S) \mapsto S$
3. Research knowledge (gaining and change of it) denoted $C : ({}^{R}S, {}^{R}S) \mapsto S$
4. Query formulation problem (QFP) as each layer of mind must know about previous layer in its own language, i.e. bottom layer must know in interaction language form the state of the perceptual layer (see IN transfer diagram, this can be denoted $C : ({}^{user}S_{cog}, {}^{user}S_{physical}) \mapsto S$)
5. A simulation design using groups of unitary operators (as in Section 4.3.7) where user behaviour changes from following a pattern $[U_i]$, to a pattern $[U_j]$[75], is said to be converging to stable behaviour $[U_j]$. This change therefore corresponds to the function $C : ([U_i], [U_j]) \mapsto S$

The SUP and QFP mentioned in the initial chapters being the key problems in IR are therefore an issue of convergence, and in addition any problems pertaining to search scenarios (employing the generic definition of the higher ostensive) involving change of knowledge (and hence statements) are also issues of convergence. The SUP corresponds to a simulated subprocess of Π that can be represented as a set of flows in the user agent $[\gamma \in {}^{User}\Gamma_{INT}]$ with $K(S_{\Gamma_{SUP}})$ denoting the user's knowledge about the system and trying to mimic the learning of a real user corresponding to ${}^{R_{trad}}S_{user}$.

For a simulated user with cognitive model that represents S_{cog} as concept combinations (i.e. by techniques in [AG02]), the query formulation problem is to measure the extent to which a query represents the corresponding concepts in S_{cog}. In general, the convergence function is necessary in R_{thesis} to relate two states that are not equal, typically (taking from R_{trad}) these two states are the user's and system's knowledge. Relative to the perception of the user and system the convergence function (in their views: $R_{system} \lfloor R_{user}$ or $R_{user} \lfloor R_{system}$) of one state's knowledge with what they know of the other state is related to the uncertainty of one's knowledge about the other.

---

[75] Recall from Section 4.3.7 that a set of evolution operators $[U_i]$ corresponds to a set of ways of changing the probabilities of relevance of items thereby simulating a relevance feedback.

If the knowledge about SUP, $K(S_{\Gamma_{SUP}})$, is gained by $[\gamma \in {}^{User}\Gamma_{INT}]$ and $K(S_{\Gamma_{QFP}})$ by $[\gamma \in {}^{User}\Gamma_{EXP}]$ then what can be said about the same type of flows in the system? The concept of user understanding problem (UUP) and visual information representation problem (VIRP)[76][77] where the knowledge $K(S_{\Gamma_{UUP}})$ is gained by $[\gamma \in {}^{System}\Gamma_{INT}]$ and $K(S_{\Gamma_{VIRP}})$ by $[\gamma \in {}^{System}\Gamma_{EXP}]$ corresponds to the problems with structures equivalent to SUP and QFP respectively. The VIRP is an HCI and software design issue already common to IR investigations as is UUP through use of user studies and information seeking investigations; R_{thesis} shows that their structures can be equated to that of SUP/QFP on a general level and that perhaps knowledge about UUP/VIRP can aid modelling of SUP/QFP for user simulation purposes[78]. From $R_{thesis}\lfloor R_{QT}$, it is understood that commutative operators in the Hilbert space can be used to model conservation properties (like energy conservation in Section 5.4.6.1). If one now considers the practical R_{CS} translation of this concept of conservation then it is proposed that this refers to comparing algorithm designs for the similar flow-semantics SUP/UUP and VIRP/QFP. Do particular operations in SUP for a search scenario require that UUP also behave in a similar way? In general, to what extent does SUP relate to UUP and VIEP to QFP? These research questions indicate a potential use of the conservation semantic of Section 5.4.6.1.

How does the oracle, which can be a simulator or in R_{trad} the user, or some component know the possible future states in order to pick from amongst them? For the flows between matching and corpus components ${}^{i}\gamma_{x}$ of R_{trad} the documents of the corpus need to be represented as statements in some language, or proxy information is used, for example, 'there exists 20 clusters 5 of which have *dog* and *cat* in common and 15 of which are about engineering'. This proxy description can also be taken to be a query to the system with informal translation '..fetch me documents with the words dog and cat in them..'; the following is the elaboration of the search process in our notation assuming a simple high-level stack structure as depicted in the example below.

---

[76] The UUP problem pertains to the equivalent of a semantic gap problem for systems as the system does not usually know many details of what user-known semantics are denoted by particular interactions, and the VIRP problem to human-computer interaction issues in $R_{CS} \prec R_{thesis}$.

[77] R_{CS} refers to the computer science perspective. This can be regarded as a highly programmatic perspective for the purposes of this thesis in the sense that the observation $R_{CS}\lfloor[x,\ldots]$ can be taken to mean a computer program of scenario $[x,\ldots]$ so that $\mathcal{L}(R_{CS})$ would be a programming language (or pseudocode).

[78] Further, this suggests a way to introduce user/system modelling issues from user studies by information seeking investigations into system modelling in R_{thesis}. These issues are suggested as open investigations.

Example 2. *[Simple Search] The process starts with* $\longrightarrow \gamma_{init} S_{int}$ *upon which the user types the query* $\Gamma_{typing} \subset \Gamma_{INT}$[79] *that is then passed to the matching layer* $S_{int} \xrightarrow{\gamma_{query}} S_{match}$ *where the query is 'dog', the matching layer then interprets this as an interest by an internal flow* $S_{match} \xrightarrow{\gamma_{record}} S_{match}$ *where* $T_{English}(\mathcal{L}(S_{\gamma_{record}})) =$*'..the word is a relevant word..'. Then from the matching layer there are a set of flows* $S_{match} \xrightarrow{[\gamma_i]} [{}^{cluster}S_i]$ *where* $T_{English}(\mathcal{L}([S_{\gamma_i}])) =$*'..i am a cluster with documents that have the word: dog, and my relevance score is...'. Corresponding to each flow from* $[\gamma_i]$ *there would be internal flows from each cluster to subclusters and then finally to the documents (if that is what is to be retrieved)* ${}^{cluster}S_i \xrightarrow{\gamma} \dots \xrightarrow{\gamma} {}^{RelDoc}S_D$[80]*. On the return there is* $[{}^{RelDoc}S_D] \xrightarrow{[\gamma]} \dots \xrightarrow{[\gamma_{relDocs}]} S_{match} \xrightarrow{\gamma_{resultInfo}} S_{int} \xrightarrow{\gamma_{show}}$ *where* $T_{English}(\mathcal{L}(S_{\gamma_{relDocs}})) =$*'..score for document 1 is .., score for document 2 is ..., ...',* $T_{English}(\mathcal{L}(S_{\gamma_{resultInfo}})) =$*'..first document in ranked list is .., second document in ranked list is..' and* $T_{English}(\mathcal{L}(S_{\gamma_{show}})) =$*'..put header of top document in bold on top left,..,place image from document on right in box, ...'.*

The difficulty in readability in this description of a simple search scenario is due to the fact that our observations are that of:

$$[R_{English} \cup R_{PseudoCode}] \lfloor R_{trad} \tag{5.2}$$

and although

$$[R_{English}, R_{PseudoCode}] \preceq R_{thesis} \tag{5.3}$$

and it shows change of state, clearly a translation is required so the perception $R_{formal} \preceq R_{thesis}$ can be used for analysis. All traditional scenario setups are roughly equivalent to the structure presented above, whether there is a complex interface that incorporates interactive intentions in [Xie02] or otherwise[81].

---

[79] Recall ${}^{int}\Gamma_{INT}$ denotes the internal flows of interface state S_{int}.

[80] There would also be internal flows in the document itself or in the index $S_{index} \subset S_{corpus}$ from the practical point of view of a R_{sim}, or also from the abstract view that sees a flow coming into a document as an instigator to multiple internal flows that affect word states that then send their response (communicate) to the higher states, this abstract view is useful for structured document retrieval as elaborated in Section 6.1.

[81] This is true also if the scenario takes into consideration other factors including an entire operating system like Apple's OS X, or a meta-filing system like Microsoft's WinFS where search notions are embedded in a stack, and this stack also contains functions of the general operating system that are not traditionally associated with search. Thus one can label the entire operational system stack (in the software engineering sense) as the search stack.

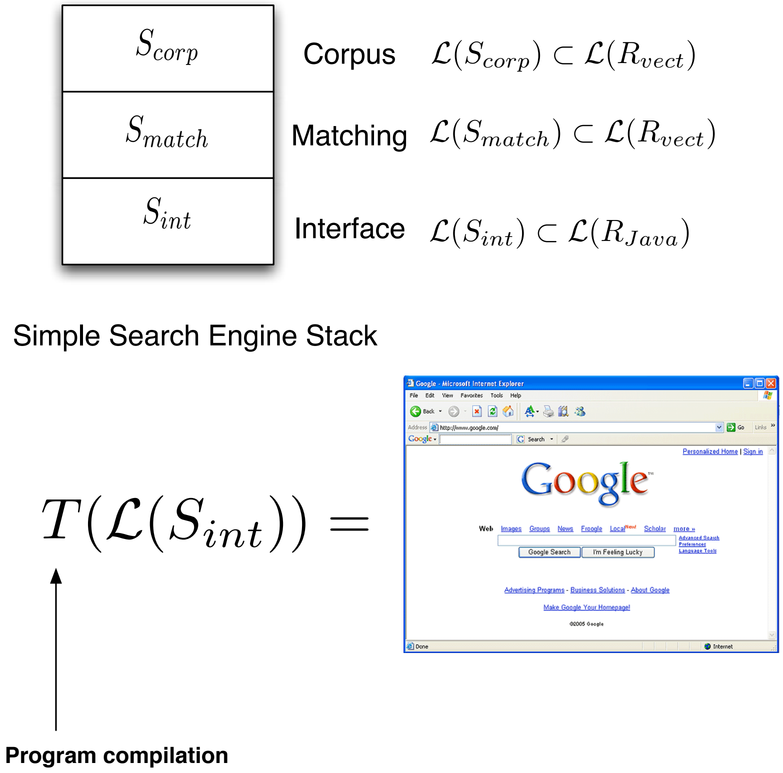


Figure 5.9. A simple search stack

This is also the case with WWOS (see Section 2.5) if it is seen as equivalent to an operating system stack. For the sake of completeness, it can also be said that for several agents, representing different machines/components/users different machines (as is familiar to the concept of a WWOS and to Web 2.0), that they could all be represented as one stack depending on how the model is to be employed. Accommodating these diverse search scenarios into the scope of R_{thesis} means that there is a potential in integrating all this for formal analysis but at the moment our notation of states, flows and their semantics are too general except for rudimentary classification (some examples and guidelines can be found in Section 6.1) as the scope allows.

5.4.8.1 Simulation and Convergence

Proceeding to the the lower level with regard to simulation, in the view of R_{trad}, functions over documents $f(D_1, D_2, .., D_n)$ are required for the system agent to be adequately simulated. Simulation according to R_{thesis} similarly requires functions over states, Fig. 5.7 shows the simulation trees[82], what is relation between the R_{trad} tree ${}_t\overline{D}_i$ and state graph ${}_tS_i$? For each ${}_tS_i = {}_tS_D \cup S \cdots \cup S_I$ where S_D is a document state and S_I is a interface, hence the relation between traditional documents and new document state is ${}_t\overline{D} = f(S_D)$, thus all functions on documents in traditional perception correspond to functions on substates of S_Π in R_{thesis}. In terms of R_{thesis}, one can ask whether or not there will ever be a search goal that can be defined in terms of non-document states such as states of components (from the stack) or even a combination of component and document states $f(S_I, S_D)$ so that the interface state is also considered. These are examples of new questions that arise naturally through our conceptualisation of search. Functions over states are essential to simulation but it is not the case that all possible next states for each component require to be modelled unless one is doing an exhaustive analysis. Therefore a component may not always know[83] all possible future states of another component with which it shares a flow, so there is uncertainty with regard to future states, and in terms of modelling one must work with approximations or samples of future states. Section 6.1 discusses the modelling of corpus and matching components and how the model of a matching stack knows what documents to get next.

5.4.9 Higher-level Semantics

During a search process the overall search state S is changed traditionally change happens ostensively in the view of researcher, system and user agents. In the simulated case, the researcher has much more knowledge about the simulation that they have designed. However, as the simulation has to take place for results to be known, the researcher may not be able to guess the final state of the search or states prior to it. Therefore, (state) change still occurs 'ostensively' in the researchers perspective, except perhaps it is 'less ostensive'[84] in the simulated case. From the researchers point of view the changes of state can be described algorithmically by considering possible future states given current and past states by means of a state picking function. This function represents the generator

---

[82] They can be graphs but assume for simplicity that the states do not repeat, and that repeating documents are different than the original occurrence due to them showing up at another time.

[83] A simulation agent may know all future states but in general an agent would not be modelled as one knowing all future states as it is not realistic.

[84] It is less so as the researcher has more knowledge about what's going on in user models in a simulation.

of a state in a search simulation, the definition of this function depends on the definition of stack entities in the agents involved.

What analysis can one do that is useful? For simulation of search scenarios, what are the primitive concepts for a simulation? Initially one can view both user and system as black boxes and reason about properties changing (by simulation) over time that may involve these agents in a general way; a simulation would end when a search goal is reached. In the visual ostensive model the user and system are referred to by the relatively ostensive change of state. For some ostensive tree T, which is essentially a state graph, the tree grows according to the following algorithm:

1. Check goal status, If search goal has not been reached *Pick* the next state of search from among available next states, otherwise stop
2. Deduce new tree T', where the change is $T \xrightarrow{a'_k} T'$[85].

Labels are a set of semantics (represented as simple English) for a process that include energy types and the *Pick* subprocess type. Hence the translation is changes $\xrightarrow[\text{syntax}]{}$ Paths $\xrightarrow[\text{semantics}]{}$ labels. It is also true that changes $\xrightarrow[\text{semantics}]{}$ labels. Further, the *Pick* is the function $f([S_{prev}], S_{new}, S_{Current})$, *Pick*$\subseteq \Pi$, *Pick* $= Paths$. A system deduces T' from T by the following the pick function schema $f([S_p], D_e, D_c)$ where $[S_p]$ are the prior states, D_e are the set of new exposed documents each corresponding to a flow (and a potential path), and D_c is the document that has been interacted with, thus user interaction is representable by flow and/or path types. A translation is an operation that changes statements thus it is also one or more flows and the word 'translation' or a label 'T' is also a semantics for flows. In particular, a translation between frameworks as observed by a researcher would refer to a set of flows in the cognition of the researcher and similarly for any function on the set of states, statements or frameworks.

A simple search goal would be to require the user agent to have interacted with a set of documents thereby denoting a type of user experience, or in traditional IR terms, a traversal of a set of relevant documents. An alternative goal, referred to as a stable

---

[85] a'_k in the traditional ostensive graph denotes the document that was selected at time k that takes T to T'. In the general ostensive graph (of states) it denotes the event or change that causes the overall state change.

state goal, given a simulated scenario with flows, is defined by a particular substate in the user that stops changing (except trivially[86]) or becomes stable even though the user is experiencing flows from the system. A stable state goal can be specified in our notation as $\forall\gamma, \xrightarrow{\gamma} S(\subset S_{cog}) = S$. A stable state goal is useful for representing several search related concepts, for example it can be used to represent that the user cognition is no longer interested in continuing the process[87]. In general, a goal is a function over the totality of knowledge at some point which is inscribed in the state of search at that point, but when all knowledge is not known then it is the knowledge known at that time, i.e. $K({}_t\Pi)$, that is used. There are $2^{|D|}$ functions of type $f([D_i]) \rightarrow [T, F], i \leq |D|$[88] denoting the set of goals pertaining to a traditional document-orientated understanding of the purpose of a search. A richer set of goals is similarly defined over a subset of all possible states of search $f(S_1, S_2, \ldots, S_k) \rightarrow [T, F], k \leq |D|$. Inherent in the above simulation algorithm is the view of search as a game[89] between the user and system as a tree evolves once the user picks their state followed by a pick in reply by the system, both combining to form the new overall search state. This state picking by each agent occurs at the top-regions of the stack, the regions representing cognitive functions. The utility function from Section 4.3.4.1, over one or more states (or documents), can be used as a search goal as shown in Section 6.4. In addition, one can use the several types of functions, as elaborated in Section 4.3.4, to form decisions about if a search goal has been reached, the particulars of this are not addressed here.

The language of flows and stacks are a 'middle language', mid-way between both R_{QT} and R_{CS} with the idea being that one can 'port' to QT for formal analysis and back to CS for implementation upon considering the findings of formal analysis. When a set of flows are grouped by a path then this indicates a set of not necessarily sequential (depending on if it is an ordered or unordered path) events and this is where dynamics are important. Also,

---

[86] A trivial change denotes a change where only the time parameter is incremented.

[87] This type of goal definition implies that after some interactions the user agent is stable, that it does not change, this appeals to formal mathematical analysis of the search process to deduce the nature of stability both empirically by statistical methods, such as the average steps required for stability, or existentially in arguing whether a particular set of interactions will ever cause stability.

[88] The documents here could also be just the top few of a ranked list as would be represented by a substate representation of S_Π.

[89] This could be a formal mathematical game or more generally a process between mutually interacting entities that stops upon satisfaction of some constraint.

$\exists^1\gamma_x \ldots {}^n\gamma_x <: \gamma_x$ that denote information passed between the general state S_{match}[90] and the general state S_{corpus}, similarly for $\overline{\gamma_x}$ the flows going in the opposite direction from corpus to matching states. So this simple model $(S_{corpus}, S_{match}, \gamma_x, \overline{\gamma}_x)$ translates well into computational languages like Java as objects and operations on objects.

Reversibility More generally, for ${}_tS_{corp} \xrightarrow{\gamma} {}_{t+1}S_{corp}$, if $\exists K(\overline{\gamma}) \in K(S_\gamma)$ then this indicates that the change γ is potentially reversible and can happen in the way a flow γ can potentially occur in a path. A path that is reversible contains all inverse flows for each flow, and can be taken to correspond to the reversing of a search decision, or of a Pick process which in turn corresponds to the semantic of opposite user behaviour, as if the user is undoing the effects of their interactions. In order to quantitatively model reversible sub-processes in search, the quantities, such as those pertaining to relevance of documents, require to be represented in a way such that there is a concept of inverse operations in the respective quantitative framework. An example of such a framework was given in Section 4.3.7. On a semantic level, $\overline{\gamma}$ would indicate that there is now a particular user behaviour that encouraged a reverse of a prior interaction and that there was *a change in direction in the IN of the user* (the utility measure of Section 4.3.4.7 quantifies this). In terms of energy it could be said that $E({}_{t_1,t_3}S_{corp})$ is conserved if the IN is invariant over all terms in this period of time, for it to be conserved there requires to be a $\overline{\gamma}$ for every γ among the interactions in which the corpus participates.

5.5 Languages and Translation

5.5.1 Introduction

In the prior sections notation was introduced by which one can describe all search scenarios. In a general way, this section addresses the issues pertaining to re-representing these general descriptions in different languages or in terms of other frameworks. Specifically, this section elaborates the translation function employed in prior sections, the language function $\mathcal{L}()$

---

[90] Which is represented as the matching component of the system in the laboratory view of IR and as a rectangle on a stack in our method.

and ordered sets $\mathcal{R}_i \subset \mathcal{R}$ of frameworks that denote all perceptions of a certain type such as $\mathcal{R}_{vect} \subset \mathcal{R}$ that represents all vector space approaches for search modelling where $R^*_{vect} \in \mathcal{R}_{vect}$ represents the optimal vector space based approach.

Recall that the function $\mathcal{L} : X \mapsto L$ that returns L the representation language of X, where X is a state, statement or framework. L is not one particular representation but refers to all known representations. As all statements about a framework are held by its state, all statements about representation of that state are also held in the state thus it is said that $S_\Pi = S_{\mathcal{L}(S_\Pi)} \cup \ldots$ without further detailing the specifics of this equality, but give the example $S_{KVR-QT} = S_{\mathcal{L}(\mathbb{H})} \cup S_{\mathcal{L}(English)} \cup \ldots$ meaning that the representation uses aspects of Hilbert space representation as in QT but discusses them in natural language. The representation of L itself or of $S_{\mathcal{L}(S_\Pi)}$ with respect to theories of linguistics is not addressed, for most cases, references to $\mathcal{L}$ can be taken to mean $T_{English}(\mathcal{L})$ where $T_{English}$ is an English language descriptor for L such as ‘vector space model’.

5.5.2 Language and Cognition

This section presents some requirements of a formal language to describe the cognition of the user agents for the purpose of user modelling. The aim here is not to suggest a particular representation of cognition but to discuss the general properties of such representations. Recall from the background that a CIR system would ideally be like a human agent from a HHIR scenario[91]. In terms of R_{thesis} this means that the research should be conducted with realistic user models and that systems should be built to satisfy these user models so that SUP and QFP are adequately addressed. The following analysis is mainly based around the reasoning of [AG02] for describing any `AU` where it is suggested that the mathematics of Hilbert space theory augmented with QT concepts form an adequate set of tools for describing cognitive phenomena.

Consider that the memory layer in a stack for a real user in the physical context on the chemical level is defined at least in part by neurons and signals where physical relations or communications between neurons can have more than one meaning. Usually cognition

---

[91] In fact, our modelling of both user and system using the same techniques is due to identifying HHIR as the goal, in this sense it can be said that the perception of R_{thesis} inherently tries to *personify* the system and make the user more system-like (i.e. by modelling it for simulation).

refers to a kind of functional abstraction of the physical space where a set of neurons and their relations are given a description that indicates their function or a *concept* they define. It is assumed that this abstraction is sufficiently general for describing ${}^{AU}S$ at this stage. Hence, descriptions of memory are assumed to be in terms of concepts. With regard to this conceptual view of cognition one would presume that ${}^{RA}S$ would be quite different, as there is no direct physical baseline from which to abstract. However, abstraction in terms of cognitive concepts can be used here without formally defining concepts. The proceeding discussion addresses some representation issues for ${}^{AU}S$, and S_{cog} in particular where it is assumed that one should be able to represent ${}^{RA}S$ in the same way as ${}^{AU}S$[92], the more realistic the ${}^{AU}S$ the more restriction in its similarity with albeit with some restriction.

Given that an abstract system `RA` is very restricted relative to a real user, its cognition by which it is meant (according to R_{trad}) its data, data representation (index or otherwise) and matching (corresponding to the 'reasoning' layer in `AU`) layers, cannot have a 'richer' description/representation than some description of a real cognition[93]. There are reasoning components in the human cognitive stack which work with and affect the memory where $S_{user} \supset S_{cog} = S_{memory} \cup S_{reasoning} \ldots$. As memory changes it is assumed that the way reasoning affects memory also changes over Π. This changing effect of reasoning on memory and representations are addressed in [AG02] which suggests a $R_{\text{[AG02]}} \prec R_{QT}$ perception for modelling cognition, thus due to generality of R_{thesis} it is said that $R_{\text{[AG02]}} \prec R_{thesis}$ such that $\exists T_1, T_2$ such that $T_1(\mathcal{L}(R_{\text{[AG02]}})) \prec T_2(\mathcal{L}(R_{thesis}))$[94]

5.5.2.1 Structure of $S_{cog} \subset S_{user} \subset S_{\Pi}$

In human cognition concepts exhibit *association* with one another and this 'association structure' is used for reasoning. Association in its simplest form is itself a concept which when specified in a language requires at least another concept to be stated with it that details the association or specifies that it has no further specification (thus is a general association). An example would be the first diagram of Fig. 5.10 which shows the reasoning between the existing associations, and can be described as the *On A* relation. Conversely,

---

[92] Keeping in line with the idea that HHIR with two human agents is optimal IR.

[93] This is suggesting that human cognition can never be completely represented in an abstract form.

[94] This is trivial as it was said that S_{cog} is a state that holds all knowledge and thus there are no restrictions to its representations so it can be $\mathcal{L}(S_{cog}) \subset \mathcal{L}\left(R_{\text{[AG02]}}\right)$.

if a description was the proposition 'Person is on a bus' then it implies association between concepts person and bus, where these relations are concepts themselves. In the conceptual structure of a S_{cog} representation for a real user, associations also have a concept of strength denoted by the line length in the second and third diagrams of Fig. 5.10, this indicates the *'closeness'* (also a concept) of the friendship. In general, if there is a statement, one can extract arbitrary concepts and relations from it. Thus concepts are abstract entities which can be defined by arbitrary length natural language statements/propositions. Usually cognition modelling is addressed by considering the processing of language which can be natural language, or as for a real `AU`, in *perceptual language*: $\mathcal{L}_P$. The latter refers to language that can describe anything experienced by a human through any *senses* creating perceptions[95].

The most powerful language would be a *'hypothetical absolute description of mind'* that is richer than the perceptual $\mathcal{L}_\infty$ and could refer to descriptions that are about non-perceptual or *abstract* things[96]. It is assumed that perceptual language is richer than natural language, which in turn is richer than the gesture language, that is $\forall \mathcal{L} \; [\mathcal{L}\,(\mathcal{L}_{gesture}) \prec \mathcal{L}_N \prec \mathcal{L}_P]$ where $\mathcal{L}_N$ and $\mathcal{L}_P$ denote natural and perceptual language respectively.

The assumed languages in this chapter are from the ordered set: $(\mathcal{L}, \prec)$ which has a partially ordered subset $\mathcal{L}_{cog-sim} \prec \mathcal{L}_N \prec \mathcal{L}_P \preceq \mathcal{L}_\infty$ where $\mathcal{L}_{cog}$ denotes the cognition description language or conceptual language that can be used to formally describe cognition whether it is the anomalous state of knowledge in [BOB82a] or potentiality of concepts in [AG02] (as indicated in Fig. 5.11); $\mathcal{L}_P$ is a hypothetical language within human cognition and the thesis does not address here how perceptions are made into concepts/relations in our minds instead it works with the particulars of a less richer and simpler set $\mathcal{L}_N$ to infer a formal conceptual language; $\mathcal{L}_\infty$ is the maximal element in the set. Missing out the perceptual language means that any derived concept formalisation that is found is relatively very ignorant of the nature of cognition as it is not derived from perceptual phenomena. Any further analysis will thus result in an inevitably weaker definition of cognition than one obtainable on considering perceptual properties. It is possible to opt for a slightly richer

---

[95] This is the cognitive science meaning of perception rather than the meaning it has held so far in the thesis as being synonymous with framework or researcher.

[96] This language could be infra-linguistic w.r.t. natural language, whether it is specifiable or realisable in some other way is a philosophical debate, and is out of scope.

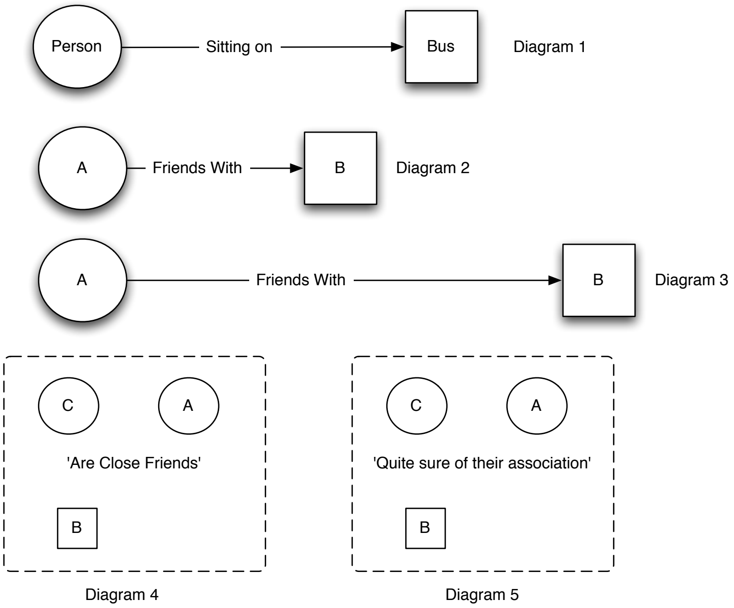


Figure 5.10. Association of Concepts

language which also includes the visual language adequate for describing images, however this thesis restricts to only analysing processing of natural language.

An information need as traditionally understood stems from the perceptual in cognition, the memory layer in the stack design, and heads towards a query, the gestural layer in stack, Fig. 5.11 illustrates this in terms of languages. In terms of our definition of information need, the IN of R_{trad} is a higher-level semantic that constitutes a combination of the IN's in some process according to R_{thesis}. The proposed framework suggests ways to exactly define this need in terms of flows and energy semantics, and measures on sets of flows can be used to quantitatively model the loss of 'information' w.r.t. R_{trad} as it is translated (w.r.t R_{thesis}) from perceptual (w.r.t R_{trad}) to the physical layer (w.r.t R_{thesis}).

5.5.2.2 Limit of Languages

Anything which can be interpreted by any of the senses forms part of language, where language is defined as that which is transmitted between two agents whether it is braille felt by touch, words recognised by vision or indeed something spoken. A language is not specific to one sense, but could require more than one sense to interpret, such as a written symbol followed by a sound emitted from the vocal apparatus. Languages may not be

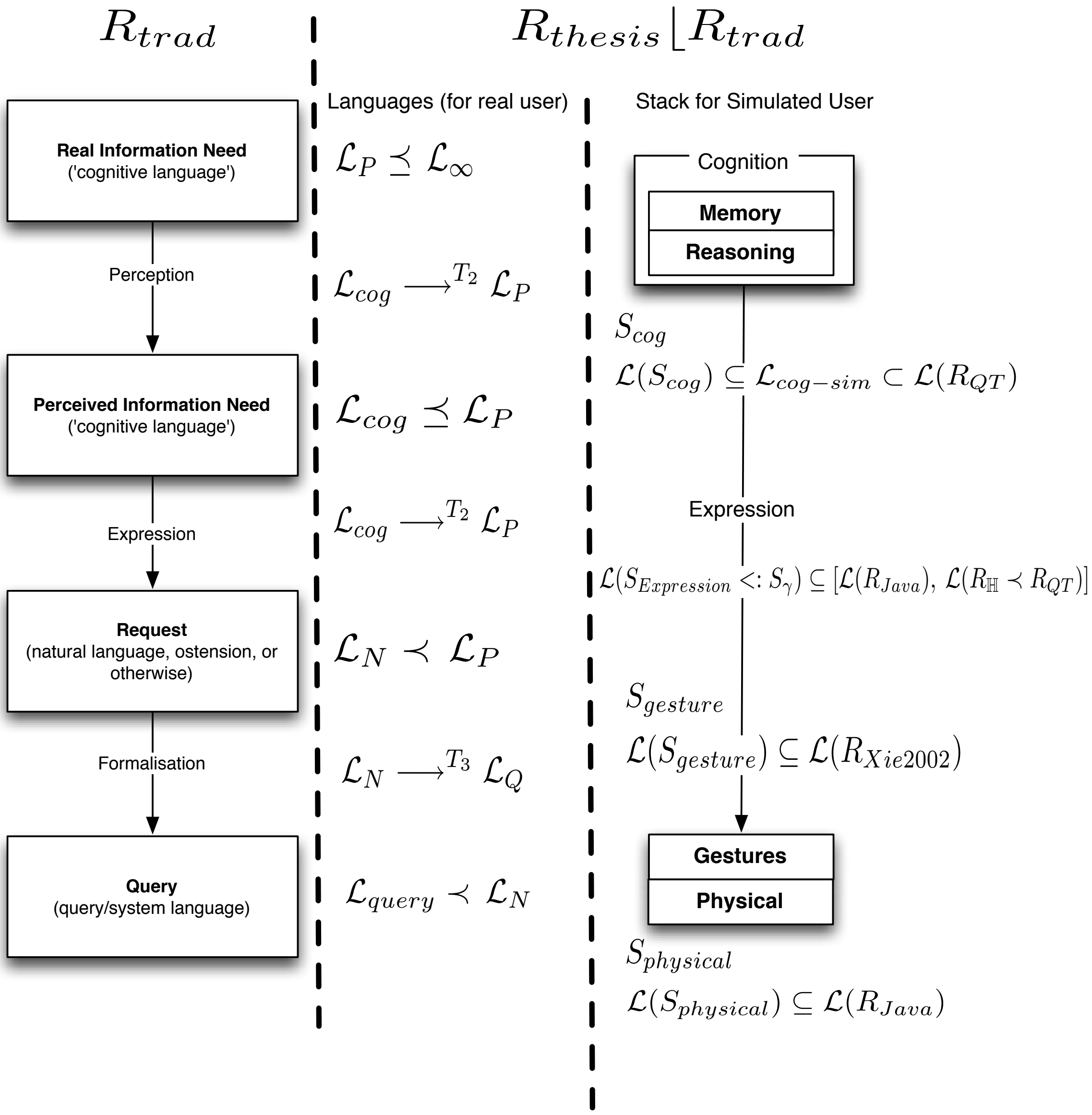


Figure 5.11. Observing traditional information need transformation in terms of languages

expressible or interpretable, for example, a particular concept in one language may not be fully expressible in another language. A mathematical equation may not be interpretable only in terms of 'taste'. When we speak of translation or other language processing occurring in an agent it is from the point of view of the creator of the agent who has a detailed view of it[97]. When the agent is human assume that a third observing entity is noting down the language processes.

Interpretation of a language corresponds to translation of it into some other form. Recall

---

[97] This view is $R_1 \lfloor R_2 \lfloor T(s_{IN_1}) \in K(S_{cog}) \longrightarrow s_{IN_2} \in K(S_{physical})$ where R_2 is a specific theory about translation of the need and R_2 is the observer using this theory.

that translation is just a higher-level semantic for a set of flows that cause statements to change in the perception that observes the translation. For example, interpretation of a mathematical equation requires a person or agent in general to read the equation upon which if a verifying agent exists then the reading agent must re-express the equation in another form to the verifier; otherwise interpretation is personal. In the first case, there are two translations from source language to thought language to re-expression language, and in the second case there is only one, the translation to thought language. As the third agent we can detail exactly what *type* of translations are occurring. Consider now a fourth agent who is similarly observing and fully understanding the third agent who is observing the second agent, i.e. $R_4 \lfloor R_3 \lfloor R_2$. If the R_4 indeed has control over R_2[98] and starts mirroring the happenings of R_2 in R_3 then R_3 is essentially thinking about itself knowing that it is doing so. This is unlike the human case when the human cannot see his mind working exactly in some detailed way. Then the question becomes, what is the nature of the processes which are self-referential in which R_3 thinks about the linguistic processes happening within it. This question may not be answerable for realistic agents when the realism is sufficiently high as then one may need to consider deep cognitive issues, as if the ideal description of a user agent is equivalent to a complete description of a human cognitive system then there are limits to the agent's description. The existence of this limit is especially suggested in the case where one attempts to create a realistic `AU` encompassing some properties of the researcher doing the modelling[99]. In an extreme version of this case where the researcher is attempting to create an agent that completely resembles himself/herself there maybe certain elements of himself/herself that they are unable to describe due to the inevitably non-objective way one perceives oneself. Out with this extreme case, any user model will be limited due to the non-objectivity of the creator of the model whether the model is depicting the creator or another human.

It is additionally assumed that the healthy human cognition does not enter a state of infinite recursion due to self-referential thinking, thus there is some sort of internal mechanism that stops one from thinking self-referentially. Similarly agents should be designed to reflect

---

[98] This is a realistic scenario if we consider R_2 to be a system, R_4 an evaluator/researcher and R_3 a simulated user. In fact the observation $R_4 \lfloor R_3 \lfloor R_2 \longrightarrow R_5 \lfloor R_4 \lfloor R_3 \lfloor R_2$ as R_5 is the perception discussing the realistic scenario, but it is a passive observer.

[99] A crude equivalent would be a automatic theorem proving system.

this assumption so that in any search scenario the agent's *continuous knowledge* of the inner workings of each other and subsequent duplication is limited. An even stronger assumption is made here, that a human cannot know the level of recursion/self-reference of which he/she is capable, thus one cannot ever fully know how any `AU` is limited in an objective sense. If a human cannot know their limit then that is a limit in itself. Thus, there is some confidence for any `AU` model in the sense that the best `AU` model can be severely deficient according to some absolute source but it is, relative to limited human cognition, something that 'just works'. Real humans do not think by self-reference at such a level[100] so it is not required for modelling, this can be said to be a natural or inevitable weakness in the model, or equivalently, it can be seen as strength as it aims to mimic reality.

5.5.2.3 Derivation of Languages

Any description language for cognition derived on the basis of considering perceptual phenomena *can* be richer than that derived considering only natural language phenomena. The richest of these is assumed to be inferior to the hypothetically richest description language for cognition. Given any mapping $D : (\mathcal{L}, \prec) \rightarrow (\mathcal{L}, \prec)$, which translates a language ordering to another language ordering preserving the order, that outputs a derived conceptual language, the following is proposed $(\mathcal{L}_G = D(\mathcal{L}_N)) \prec D(\mathcal{L}_P) \prec \mathcal{L}_\infty$ where $\mathcal{L}_G$ is a model of cognition. Hence there are no mappings that can improve a language to a 'higher class'.

5.5.2.4 Cognition Modelling issues

The strength or degree of association can be seen as a *property* of the association or concepts themselves and be used to *generate* or *classify* further associations as depicted in the fourth diagram of Fig. 5.10. There is also the notion of uncertainty attributed to associations that differs from strength as it is an indication of certainty with respect to both the association and its strength. Uncertainty can be seen as a property of association or an association itself, attaching a value to it would allow us to talk about the *degree*

---

[100] Except perhaps in extreme cases.

of uncertainty, this in turn can be used to classify concepts so that all those concepts involved in associations of the same degree form a class. Classification with respect to uncertainty of some degree is a newly *generated* concept, as shown in the fifth diagram of Fig. 5.10. Indeed there are many concepts similar to simple (weak) association and association augmented with strength or uncertainty. One of these concepts can be in turn used to classify others in the same way the proposition ‘people taller than 6ft’ can classify a general set of people or generate a new set of concepts that satisfy the proposition. These associations between concepts can therefore form classes each of which can be represented by some real number denoting the degree of association.

An initial description language structure fitting these conditions can be defined as

$$\mathcal{L}_G \supseteq \left[\left(\Sigma_{\text{Concepts}}, [\text{Associations}]\right), \ldots\right] \tag{5.4}$$

where an example model for associations is:

$$\text{Association} = \left(\alpha \subseteq \sum_{Concepts}, y \subseteq \Sigma_{\text{Concepts}}, k \in \mathbf{R}\right) \tag{5.5}$$

which is a triple consisting of a set of concepts that describe the meaning of the association, the associated concepts, in the class and finally a real value to indicate degree of membership to the α. This formalisation however seems too general and is unclear as to how exactly the concepts are associated. For example it is not obvious how to represent the proposition ‘very strong people and very tall people’ where the ‘very’ would correspond to a real number, the strong/tall to a class descriptor type $a \in \sum_{Concepts}$ and ‘people’ referring to class members. In a retrieval sense, the user’s IN could be represented by the phrase ‘really interested beaches and slightly interested in cars’. Any natural language statement can be expressed in such a language, whether a particular model is more appropriate than another for some Π is not addressed in this thesis.

Using $\mathbf{R_{thesis}}$ to model $\mathbf{S_{cog}}$ A R_{thesis} orientated model is $\mathcal{L}_G \supseteq [(S_{cog} = (S_{\gamma\in\Gamma} \subset S_\Gamma) \cup S_{concepts} \cup \ldots), \ldots]$ where S_Γ denotes the flows whose general semantic of potential to change corresponds to the concept of association and $S_{\gamma\in\Gamma}$ would hold the mean-

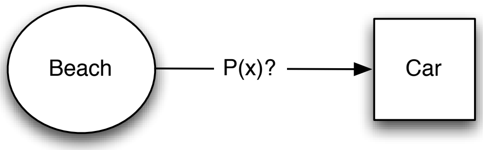


Figure 5.12. Uncertain logic from associations

ing of the association in terms of a concept such that $S_{\gamma \in \Gamma} \supset S <: S_{concepts}$. Hence, R_{thesis} can accommodate potentially rich cognitive models, such that it can be said that $R_{thesis} \lfloor R_{[AG02]} \lfloor S_{cog}$[101].

5.5.3 Translating between Frameworks

5.5.3.1 Introduction

This section discusses translation of frameworks, in particular translation between the traditional models and into QT and associated problems, aiming to show why it could be more effective to build from scratch while keeping within the conceptualisation of R_{thesis}.

Analysis of state change of a search is isomorphic to analysis of change of statements in that search. The problem with IR is that for some s_i and s_j, $\mathcal{L}(s_i) \subset \mathcal{L}(R_{vect})$ and $\mathcal{L}(s_j) \subset \mathcal{L}(R_{logic})$ and also $\exists T(s_i \longrightarrow s_j)$ but this T is not generally known, also it could be that according to some logical perception $T(s_i) \rightarrow T(s_j)$ or $T(s_i) \Leftrightarrow T(s_j)$ or $T(s_i) \vdash T(s_j)$ but due to different languages used to specify these statements one is unable to easily find the translations that exposes these relationships[102].

5.5.3.2 Observation ($\lfloor$) and translation

If a researcher is observing a scenario with a system where s_a is a statement pertaining to system and ${}_t s_R$ is one pertaining to the researcher, that is $\lfloor [s_a, R_{thesis} \lfloor s_a]$, then if at a future time the statement is transformed to ${}_{T>t} s_R = [s_{R_1}, s_{R_2}, s_{R_3}]$ such that $s_{R_1} =$'$\ldots s_i \ldots$'

---

[101] Whatever cognitive model is representable by QT representations can be understood in terms of R_{thesis}.

[102] Statements in the language of the logical, probabilistic or vector space approaches may not be known by a researcher due to them being unfamiliar with these models, instead they may just know search facts.

then it is said that observer/researcher has rephrased or translated s_a into three different statements according to three not necessarily differing perceptions. Thus, over time statements are transformed or translated within a framework with respect to the researcher and also between frameworks with respect to researcher knowing multiple frameworks, all in the eyes of the observing researcher. A framework aims to 'rephrase' or 'translate' s_i perhaps in terms of a theory inscribed in framework R_2, that is if $R_1 = f(T_{R_2}(s_i))$[103] where f is a map, then $s_L \in K(R_{logical})$, $\mathcal{L}(s_V \in K(R_{vect})$, $s_P \in K(R_{prob})$, and $s_Q \in K(R_{QT})$. Such that

$$R_{KVR-QT} \lfloor [R_{logical}, R_{vect}, R_{prob}] \lfloor s_A \tag{5.6}$$

thus $R_{QT} \lfloor [R_{KVR-QT}, \lfloor s_A]$[104].

In order to do devise completely quantum models one must be able to approach search phenomena logically, geometrically and probabilistically all at once, so one must know ${}^{Q}R_1 = f(T_{R_2}(s_i))$ where $\mathcal{L}(R_2)$="QT" and T is a full translation into QT methods instead of the partial translation in R_{thesis}[105] or R_{KVR-QT}. So if researcher knows s_v and some others, it could be the case that given s_v and others that $\vdash s_L$ i.e. a logical statement could be derivable from a vector space framework, but how is this done? This difficult problem can be approached by the translation $T_{QT}(s_v \cup \ldots)$ where $T_{QT} : s_L, s_V, s_P$ then $T_a(\ldots) \vdash T_{QT}(s_L)$ and this could be easy to show. Problem is that each s_V contains several references, for example s_{V_0} ='these documents are clustered[106] with threshold 0.7' where 'document' could refer to s_{V_1}'cluster=space, doc=vectors, ...', and threshold could refer to s_{V_2} ='threshold is the value of v_1 and v_2 dot producted' which itself refers to sV_i ='dot product is ...'. The value 0.7 could refer to s_{V_3} ='0.7 means...'. A researcher does not necessarily know all of these statements or know it consciously, the majority of this is found by investigation.

An important implication of the above discussion is that one must build from particulars, so that everything known about a process is expressed in terms of flows/stacks. Matching

---

[103] The translation function T_x is one where x indicates in English the descriptor of the destination language.

[104] It could be useful to determine if observations can be reduced, for example when is it the case that $\lfloor\lfloor\lfloor \leftrightarrow \lfloor\lfloor$?

[105] Note that the modelling in Chapter 4 was also only a partial translation due to the 'custom use' of the QT framework.

[106] Assuming a cluster representation in terms of spaces/subspaces and other geometric structures.

statements from differing frameworks by usual research is difficult, due to ambiguities, the reason for which the definition problem exists, and why search is not science. It is proposed that R_{thesis} can be taken as base language such that one can build search models from the primitives of R_{thesis}.

Let the user need be represented by the statement s_u =‘car and cats’, as documents are set of words how can s_u be associated with statements about user interaction s_t? First one needs to translate such that $T(s_u) \in \mathcal{L}(R_{vect})$ and $T(s_v) \in \mathcal{L}(R_{vect})$. Then, one needs to translate the concept of user satisfaction to say what satisfied means in adequate terms: $R_{QT} \longrightarrow^T \ldots \longrightarrow^T R_{thesis} \longrightarrow \ldots \longrightarrow^T R_{QT}$[107]. Translation of statements to stacks and flows is done by human research, which is not covered in this thesis except briefly in Section 6.1, instead actual models are addressed. If s_1 =‘user has need of car documents’, s_2 =‘D_1 =hello, dog, D_2 =john, smith’ and s_3 =‘is document D_1 mentioned in s_2 adequate for the user in s_1?’, these statements represent a research problem which itself is a search scenario denoted $R\lfloor[s_1, s_2, s_3]$ where the researcher R is observing through the abstract device of their cognitive system. A higher level view is $R_h\lfloor\ldots\lfloor R_i\lfloor[s_j]$ where some unknown statements $[s_x] \in K(R_h), R_h\lfloor[\ldots]$ are in fact the answer to s_3 above. A researcher's function is therefore to translate, or depending on the research ontology[108], to find the translation of s_1, s_2, s_3 such that $s_i \longrightarrow T_{R_{vect}}(s_i)$ (a vector space translation)[109]. A research problem can be expressed by statements at a point in time for example $[s_3, s_4] \in K(R)$ meaning the researcher will know these research statements. Over time research changes the statements into new ones or *finds* new statements, if $K(R(_0\Pi)) = [s_i], K(R(_1\Pi)) = {}_0[_0s_i \cup {}_1s_i]$ then $\exists s_x \in K(R(_0\Pi)), s_x \in {}_1s_x, \exists s_y K(R(_1\Pi))$ such that $s_x \overrightarrow{T}_{11} s_y$ or $s_y \overrightarrow{T}_{21} s_x$, where T_1, T_2 are ways to make new statements from old ones and relate newly appearing statements with old ones respectively. These ways must also be part of the researchers knowledge so it is said that $(K(f_1(T_1)), K(f_2(T_2))) \subset K(R(\Pi))$[110], thus if $T \subseteq [_0s_i \cup {}_1s_i]$ then the three entity relation (s_y, T, s_x) that links two statements

---

[107] Mapping to quantum theory has the additional advantage that the translation $\mathcal{L}(R_{QT}) \longrightarrow^T \mathcal{L}(R_{QT-computer})$ is easier and thus hypothetically one could benefit from increased computation power, useful especially for simulating complex models i.e. of cognition, interaction possibilities and corpora.

[108] The ontology includes a purpose, but recall that as re-search is search and search ontology has been discussed in prior sections and chapters, as therefore has research ontology.

[109] This is a very crude way to indicate a computational philosophy of science.

[110] So either T_1 or T_2 or some function of these maps (some alternative form) is in the researcher's knowledge.

by a third set of statements holds[111].

If the translation is not known by the researcher, or the researcher or observer observing the researcher does not know if the researcher knows (as perhaps the researcher decided to associate s_y to s_x intuitively without explicit knowledge of the translation), then it could be that the higher level researcher or research oracle $R_{orac}\lfloor R$[112] knows, and eventually at the final time ${}_{t=final}s_i$ when the research process Π has terminated there could be statements $[s_{\beta_i}] \subset K(R_{orac})$ due to observation $R_{oracle}\lfloor[[user, sys, \dots], [R\lfloor[user, sys, \dots]]]$ that indicate that the particular translation has taken place[113]. This means that the observer deduced that the researcher performed the translation to relate statements within the particular search process.

5.5.3.3 Difficulties with R_{KVR-QT}

What is hard to do with traditional research in terms of $R_{KVR-QT}\lfloor R_{trad}\lfloor \dots$ is that it is unclear how QT semantics can be used for IR and in turn when and how to use formal structures from QT in a consistent way. Any vector space perception refers to other perceptions such that $(R_{vect} \cap [R_{trad} \setminus R_{vect}]) \neq \oslash$. This is also the case for the other classical models R_{prob} and R_{logic}. In order to make a comparison between them, that is to deduce by the perception of R_{thesis} which approach is best for a particular process, one needs to be able to express them all in some form by observing $R_{thesis}\lfloor R_{vect}$ so that one can express them in terms of our conceptualisation, Thus one needs to find the maps such that $R_{vect_i} = f_i(R_{thesis}), R_{prob_j} = f_j(R_{thesis}), R_{logic} = f_k(R_{thesis})$ and if enough of these $[f_i, f_j, f_k]$ can be deduced then one can attempt to characterise the optimals $R^*_{vect}, R^*_{logic}, R^*_{prob}$. Note that $\mathcal{L}(R_{thesis})$ is semantically similar to $\mathcal{L}(R_{Dirac} \subset R_{QT})$ as it offers high-level view of search which is represented in such a way as to embed the concepts discussed in Chapter 3 in an 'automatic way' as Dirac notation does for the mathematics of $\mathcal{H}$.

Non-simulated research can be denoted by $R_A\lfloor[user, sys]$ where if R_A denotes the informa-

---

[111] This is equivalent to the concept of association in Section 5.5.2.4. It is an open question whether for a triple $(a <: s \in \Sigma, b <: T <: \mathcal{FT}, c <: s \in \Sigma)$ $a = b$ or $c = b$, that is, whether theory b (or T in example) relating a and c can generate itself.

[112] That is the modeller of the research scenario.

[113] Such that $[s_{\beta_i}] \subseteq K(S_T)$.

tion seeking perception then $['\text{intended}','\text{not intended}',\ldots] \subset K(R_A)$ and to amalgamate the research together one requires $R_A \longrightarrow^T R$ which is difficult. For $R_B = R_{QT_{hypothetical}}$, $R_C = R_{KVR-QT}$ one needs $R_C \longrightarrow R_B$ or $R \longrightarrow R_B$ which is hard and clearly one cannot translate arbitrary statements to R_B, this difficult as QT representation is a totally different medium need syntactic and semantic maps. The translation of R_i to R_A (arbitrary perception to a information seeking perception) is indirectly happening in researchers mind but this is a map of that to clarify things. Instead, one requires to start from scratch with a unified language so that $[R_1,\ldots,R_n] \longrightarrow R_T \longrightarrow [\sim R_1,\ldots,\sim R_n]$ where the middle researcher is a middle language that can be easily translated to and from.

What conditions should R_T fulfil? It should be able to encompass any and all search as defined in previous chapter (i.e. all processes), traditional search should map on easily as should user modelling and in general it should be a complete and comfortable language. Like Hilbert space theory for QT, R_{thesis} needs a particular usage technique. If one opts for this higher ostension or state picking then level of granularity, using R_{thesis} in the way of state picking, let $[u_1,u_2,u_3],[s_1,s_2]$ denote possible future states for simulated user and system respectively. Pick s_1 or s_2 then pick one of $[u_1,u_2,u_3]$ then perform analysis and see if search goal is met[114].

5.5.3.4 Simulation Techniques

There are three increasingly detailed ways of looking at agents for simulating a search involving them. At the initial level is the black-box perception of agents which denotes how agents communicate with each other at the external level (i.e. in terms of external flows). The ‘white-box’ method adds more detail to the agents by including the details of components (their internal details), and finally one can add a final dimension of detail by elaborating the search goals which dictate how and when the simulation process would halt. The first two levels are interpretable as a visual system in terms of the ostensive graph as the graph either denotes the pre-interaction and post interaction states of the system, or it is more detailed and shows the internal states for each component. Given a tree of possible future states figure out which one to:

---

[114] Discussion about translation of this into a practical scenario is out of scope.

1. Pick (a new document), if search goal has not been reached (check 4)
2. Deduce new tree which results from the picking
3. Go to 1
4. Check goal[115]
    (a) HaveSeen($D_1, \ldots, D_k$)[116]
    (b) $f_1(D_1, \ldots, D_k)$=True, and how many such functions $2^{|D|}$?[117]
    (c) $f_2(S_1, \ldots, S_k)$=True, is a more general function such that $S_{f_1} <: S_{f_2}$[118]
    (d) f_3(current state) $\subset f(S_1, \ldots, S_N)$ i.e. does some fact about this state agree with some calculation done on other states that have passed (or expected future states)

The above algorithm is viewing the user and system as two players in a state picking game[119]. In the first step of the above algorithm note that

$$Pick = [_{t_1<t}[S_{prev}], [\text{New Exposed States}], S_{t\Pi}] \tag{5.7}$$

is a subprocess, thus $Pick \subset \Pi$. $Pick$ is also a higher-level semantic for a set of flows/paths, so $Pick = [Paths]$. The word $Pick$ is a label$\in$ [labels] $\subset \mathcal{L}(R_{thesis})$.

Labels are a set of semantics for a process in terms of Fig. 5.6. [changes] $\xrightarrow[\text{syntax}]{} [P] \xrightarrow[\text{semantics}]{}$ [labels], also [changes] $\xrightarrow[\text{semantics}]{}$ [labels]. The energy type system is a labelling of paths as it denotes semantics of change. The nature of these f's are such that

$$f([\gamma_k]) = [[S_{prev}], [\text{New Exposed States}], S_{current}] \tag{5.8}$$

where γ_k denotes user interactions (and/or other changes), but in the first case, f is related to search goal, if one lets the goal $= f([S_{prev}])$ and $T_{English}(f)$ =‘to experience

---

[115] It is proposed that most search goals can be expressed in the form of these four functions.

[116] The k here could be top the k documents or some other subset.

[117] Any heuristics about this goal would be inscribed in a state S_{f_1}.

[118] Finding out whether $S_{f_1} \preceq S_{f_2}$ or $S_{f_1} \subset S_{f_2}$ would be useful.

[119] Here it is noted that it is state picking and not document picking. So this is a generalised notion that includes R_{trad} approaches to modelling interaction.

certain states' then a goal can be described a set of states $[S_g] <: S_\Pi$. For example $S_{g_1} \in [S_g], S_{g_1} =$ [cats $\rightarrow$ movie] with $T_{English}(S_{g_1}) =$ 'after seeing the document on cats I want to see movie information', similarly $T_{English}(S_{g_2}) =$'reading a document on physics' and $T_{English}(S_{g_3}) =$'dragging relevant documents into a box'. The user clicking a topic title then dragging the topic icon into a box corresponds to augmenting the knowledge of that topic title state onto all future states of search. In summary it is said that R_{thesis} is closer to R^* than R_{KVR-QT} in terms of order on $\mathcal{R}$ but also in the sense that it is easier to derive search frameworks from R_{thesis} than it is from R_{KVR-QT} (or R_{QT}).

5.6 Summary

This chapter develops the conceptualisation of search from Chapter 3 according to the state based modelling techniques of QT as elaborated in Chapter 4. A researcher's perception of search was defined as a set of statements, and relationships between perceptions pertain to relations between the sets of statements that are knowable to these perceptions.

The static information about a search is completely embedded in the search states that is composed of a hierarchy of sub-states representing the different agents and components in a search. Flows between states were introduced for specifying arbitrary static relationships between agents and components in a search. Flows also denote *potential dynamic* relationships between objects, that become actual dynamic relations with respect to a path which is a sub-process of a search process. This chapter introduced methods for assigning semantics to flows, paths, and states, some of which were borrowed from physics such as energy and related to corresponding retrieval semantics such as information need. This chapter briefly discussed representing cognitive modelling in the proposed notation, issues regarding translating the notation to corresponding notation in QT and CS; and discussed relationships with [vR04]. Further, it discussed the difficulty of translating R_{trad} and also showed how the concepts of search and re-search, as related in Chapter 3 can be related by the methods and in terms of the notation presented in this chapter.

This chapter presented a general method for generating simulations of searches by the technique of (ostensively) *picking* future states according to the higher ostensive principle of Chapter 3. It related simulation goals to the idea of convergence, shared-knowledge of agents, and the SUP and QFP problems as elaborated in Chapter 2.

—The only true wisdom is in knowing you know nothing

Socrates (470-399)

6

Applications and Discussion

6.1 Introduction

In this chapter I apply the concepts of the prior chapter to discuss some retrieval issues pertaining to structured document retrieval, user modelling and evaluation methods to show the ability of our notation to fluently represent ideas which are difficult to effectively express in R_{trad}. The idea is to show that due to the structure of R_{thesis}, agents with structure are easily represented, thus at the least, one can define them, and then to elaborate on the methods of creating usable models from these descriptions of which only a few are specified since many of these techniques are clear from the outset.

6.2 Structured Document modelling and retrieval

A substate structure for documents following a rigid hierarchy could take the form $S_{word} \subset S_{sections} \subset \cdots \subset S_{doc} \subset S_{corpus} \subset S_{system} \subset S_{\Pi}$, or the form $S_{word} \subset \ldots \subset S_{topics} \subset \cdots \subset S_{system} \subset S_{\Pi}$ or a combination of forms $S_{system} = (S_{word} \subset S_{topics} \subset \cdots \subset S_{system_a}) \cup (S_{word} \subset S_{sentences} \subset \cdots \subset S_{system_b}$. The relation $K(S_{sec_1}) \subset K(S_{sec_2})$ implies that the content of section one *refers to* the content of section two, that there is a semantic relationship between them, or it could be said that the content in one section *knows about* that in the other section. In document states, flows would be used to tackle complex relations[1]. Intra-document relations are depicted using flows in Fig. 6.1 where the document state is represented by three substructures. Consider the six element path $P \in Paths{:}P = (S_{word_1}, S_{word_{k-4}}, \ldots, S_{word_k})$[2] which denotes a change scenario that can be used to represent the effect of change in some statement ${}^{rel}s_i$ that is known by each word state S_{word_i}, thus $[s_j] \subset K([S_{word_{[1,k-4,\ldots k]}}])$ and there are translations $\hat{s}_i = T_{update(i)}({}^{rel}s_i)$ where $T_{update(i)}$ corresponds to P. An example of P from R_{trad} can be illustrated if one does a translation into technical English such that $T_{technical}(\hat{s}_i) =$'Relevance Value=0.6' and then $\forall T_{update(i)} T_{technical}(T_{update(i)}({}^{rel}s_i) = T_{technical}(\hat{s}_i) =$ 'Rel value = previous value + increment ...' denotes that the effect of P was to update relevance values. The path P is then a static set of changes, that can be given the semantic of a 'change rule' to be

---

[1] There is no need for substates to reflect this

[2] Thus the first word causes change in the *(k-4)*th word from which changes are adjacent up to k.

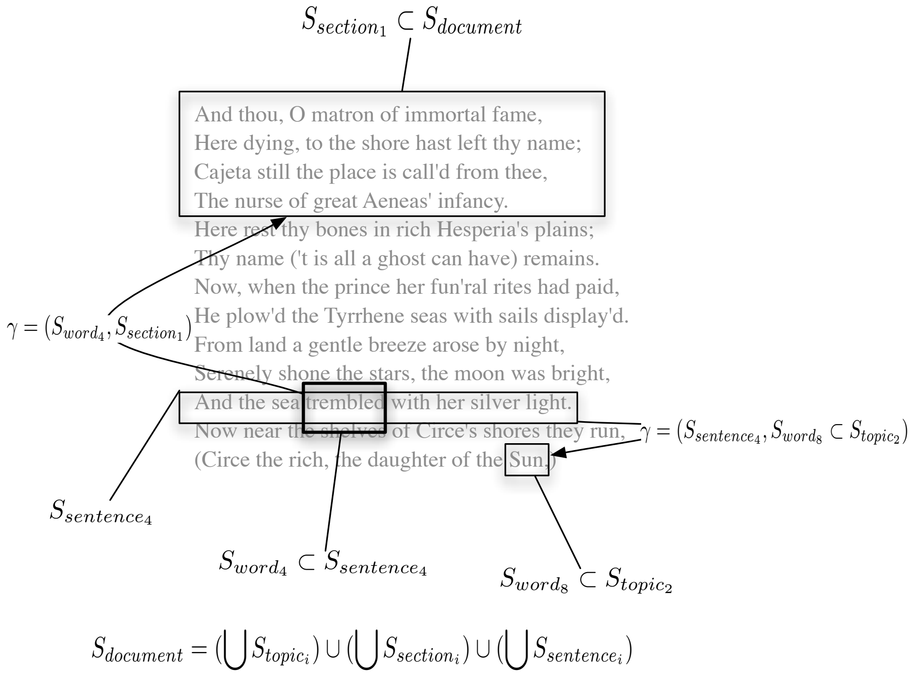


Figure 6.1. A document structured in three ways, by sentences, sections and topics, with flows between items

initiated upon a particular user interaction.

The translation $T_{update(i)}$ or an equivalent one that operates on a particular representation of relevance values in ${}^{rel}s_i$ for words can be represented by a mathematical framework and this can be used for deterministic mathematical analysis. State structure can directly reflect document structure, for example, $S_{document} \supset S_{section} = S_{title} \cup S_{text}$ where the structure of states could directly follow the document structure $S_{text} = S_{sec_1} \cup S_{sec_2} \cup S_{sec_3} \dots$ all the way down to the word or letter states. It is useful to talk about flows between states at different levels, for example, if the user views a particular section in one document it could mean that particular topics or sentences in other documents could be important, thus if one includes the flow $S_{word} \subset \cdots \subset S_{doc_2} \xrightarrow{\gamma} S_{sec_1} \subset S_{doc_5}$ in our design of the system as depicted in Fig. 6.2, evaluatory measures $[S_{measure_i}] \subset^* S_{eval} \subset S_{\Pi}$ or simulation software $T_{Java}\left(\mathcal{L}(S_{algorithm} \subset S_{sim} \subset S_{\Pi})\right)$ would require to include it in their states in order to take account of the way change should be ‘distributed’. A traditional retrieval in R_{thesis} is a pair of paths $\Pi \supset \pi_r = (P_{user}, P_{system})$ one belonging to a `AU` and the other to `RA` ; in the context of R_{thesis}, it is not a document that is ‘retrieved’ but a state which could hold details about documents. This implies that the distribution of change, updating relevance values and such, are all retrieval processes.

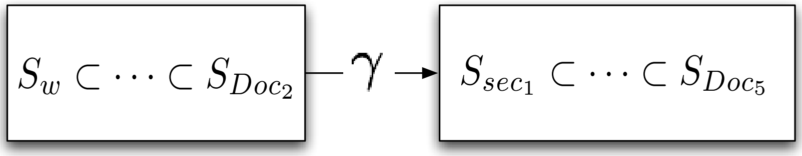


Figure 6.2. A flow between elements at different semantic levels, a word and a section in a document in this case

Modelling of intra-document (and also inter document) relations is difficult in R_{trad} especially within the canonical models. The R_{thesis} framework allows modelling of complex inter/intra-document relations in the same way as it models relations between higher structures like agents, this allows comparison between the structures of agents in a search and documents (and other components). One expected advantage is that this way of modelling can give insights into semantics in the way the R_{KVR-QT} conceptualisation gave retrieval insights. If one particular component in a system stack, namely the one representing S_{corp} can exhibit rich structure for modelling structured documents then what about the other components? What would it mean for S_{int} to exhibit similar structure? One application where S_{int} would exhibit rich structure is when the system is a question answering system that returns structured answers or narratives in natural language, that are deduced by a narrative processing component[3] $S_{narrative-processor} \subset [S_{matching}, S_{int}]$ instead of ranked output of simple graphics like that on the ostensive interface[4]. This application and specifically $S_{narrative-processor}$ would require to address the UUP and VIRP for the narratives to be useful to the users.

The discussion below considers issues pertaining to translating the structured document model state structure given above into R_{QT} for formal analysis. For simulation, one needs to know how to translate this into a simulation language. A simulation language is a language that allows specification of algorithms that can hypothetically be run on a machine, this machine is then said to be a ‘speed oracle’, meaning that it can through processing of algorithms obtain knowledge that would take the pen/paper method much longer, thus the machine knows before the pen-paper method relative to which it is an

---

[3] Such a component would intelligently structure data from result documents into a journalistic form.

[4] On a more general level if the research in academic HHIR is said to be addressed by R_{thesis}, then a structured answer/interface can be used to denote a research paper where the agent is a human researcher and the query is a research problem.

oracle with respect to time[5].

6.3 Translating to QT and Conforming to R_{QT}

The structures available to the QT framework seem to warrant an attempt at modelling structured document retrieval in $\mathcal{H}$ and translating into QT would add useful tools for analysis. There can be representational benefits by simply mapping document substates to subspaces in $\mathcal{H}$, but in order for the representation to be of use it must conform to the way QT uses $\mathcal{H}$. What is required is to translate the structured document states of the previous section into a language of QT so that it **conforms** with QT semantics[6]. How does one then model conformance? This is approached by a *meta-model*[7] of all structured document retrieval using the stack design in Fig. 6.3 where the system consists of an interface S_{int}, a corpus S_{corp}, and matching layer S_{match},. The user consists of his/her interactions also referred to as the interface state/layer $S_{u.int}$, and a cognitive layer S_{cog}, which decides particulars precluding an interaction. First the user decides something S_{cog} then the flow $S_{cog} \xrightarrow{\gamma_j} \ldots \longrightarrow S_{u.int}$ happens containing the results of the user's decision as a parameter. This causes changes in the system $S_{u.int} \xrightarrow{\gamma_e} \ldots \longrightarrow S_{int}$ which leads to retrieval of documents $S_{int} \xrightarrow{\gamma_1} S_{match} \xrightarrow{\gamma_2} S_{corp} \xrightarrow{\gamma_3} S_{match}$ that in turn changes a set of substates of the system's interface by way of internal flows $[\gamma_i] \subset \Gamma_{S_{int}}$ that cause the changes $[_j S_1 \subset {_j}^{\Pi} S_{int}]$ according to ${_0}S_1 \xrightarrow{\gamma_1} {_1}S_1 \xrightarrow{\gamma_2} {_2}S_1 \ldots \xrightarrow{\gamma_{|[\gamma_i]|}} S_1$ for $|[\gamma_i]|$, these internal changes could refer to re-organisation of icons such as on the ostensive interface in [Cam00].

Once user interactions are captured the flow $S_{int} \xrightarrow{\gamma_1} S_{match}$ causes statements to change in S_{match} that must include a mapping between interface changes and documents, in R_{trad} this mapping is between documents (represented in interface) or terms and index items that represent the corpus, one can represent this by $S_{match} = [S_{word_i}] \cup [S_{word_j}] \cup [S_{icon_k}] \cup \Gamma_{word,icon} \cup \Gamma_{document,icon} \ldots$ where the internal flows specify a static mapping between

---

[5] Similarly in complexity theory the theoretical agent Alice is an oracle according to R_{thesis} to the theoretical agent Bob as Alice is able to obtain answers that Bob may only be able to attain at some later time.

[6] Recall the issue of modelling state change in QT in the quantum chapter.

[7] That is, a generic method or modelling schemata from which specific models of document can be made.

interface elements, words and documents. Note that a state in R_{thesis} would correspond to a state vector and corresponding state-space in R_{QT} and the unions of states could correspond to several ways to combine 'sub-systems' in R_{QT}. One such way is by using tensor products of $\mathcal{H}$ spaces to indicate the different types of the states[8], this would employ the fourth postulate of quantum mechanics pertaining to combining multiple systems[9].

Possible languages for further specification of the layers can be an object-orientated language for the display: $\mathcal{L}(S_{int})$ =Java, a set of mapping functions for S_{match} which in CS-language would be a data structure, or in QT's mathematical language $\mathcal{L}(S_{match})$ is Mathematics and a Hilbert space of vectors representing documents for the corpus: $\mathcal{L}(S_{corp}) \subset \mathcal{L}(R_{QT})$. More importantly, what is required is for the way (details) of $S_a \longrightarrow S_b$ i.e. that of a state affecting another, to be evaluated according to if a user favours it or not.

Recalling the definition that re-search is search, and taking the event of looking at research paper in order to extract assumptions (of the approach described in the paper); how can one model this? It depends on the representation of S_{doc}, if one assumes $S_{doc} = S_{Assumption} \cup \ldots$, that the assumptions are identified, then this is the trivial case, otherwise a set of flows would require to be devised that exploit internal document flows that indicate semantical relationships indicating when a sentence is a possible assumption. When the flow $S_{doc} \longrightarrow S_{corp}$ comes, some internal set of flows $[\gamma_{AssumptionExtract}]$ work on S_{doc} to extract assumptions so that if one translates $T_{\mathcal{H}}\mathcal{L}(S_{doc})$ then the translated flows are such that $\mathcal{L}([\gamma_{AssumptionExtract})]$ could be represented by projectors[10].

In representing document structure using QT formalism one has to represent S_{corp} and the map $S_{match} = [S_{word_i}] \cup [S_{word_j}] \cup [S_{icon_k}] \cup \Gamma_{word,icon} \cup \Gamma_{document,icon} \ldots$. The former is shown in R_{KVR-QT}, for the latter, transformations are needed on $\mathcal{H}$. In general, finding sets of transformations to represent flows between words/icons or documents/icons (all represented as subspaces) is a non-trivial problem[11], especially if one wants to then use

---

[8] Where a particular type of state such as the interface would be considered to be a completely different system to another search component like a matching component.

[9] This would be an opportunity to employ the theoretical tools of quantum histories [Gri02] to model change over several interacting physical systems corresponding to sub-components of search.

[10] The projector would have information about what an assumption is so that when it is applied to a subspace representing a document or part of document it can extract, through matrix multiplication, the subspace of the document representing the assumption.

[11] This is due to the mathematical rules that must be obeyed to which the map in S_{match} does not

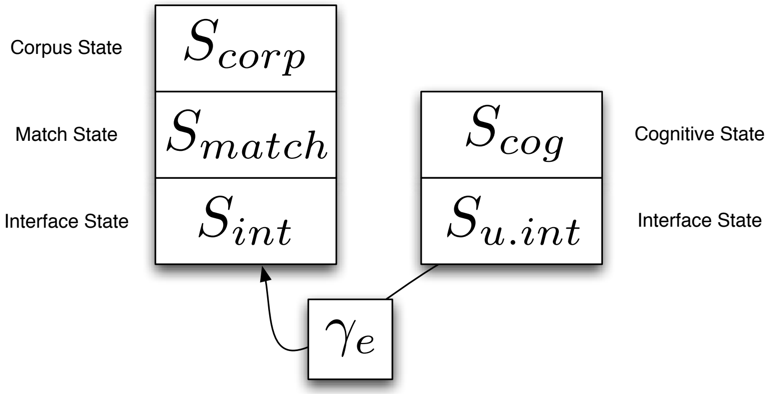


Figure 6.3. A simple stack design

the representation to model change over time perhaps by operational methods from QT, as then, more conditions are required to be met by the transformations[12]. Recall that the most general language in all our analyses is technical English such that $S_{\mathcal{L}(R_{thesis})} = S_{\mathcal{L}(R_{English})} \cup \ldots$.

Conformance can be discussed by our notation, for example, document representation in R_{KVR-QT} is represented by a vector in $\mathcal{H}$ whose bases are term vectors, a convex combination of document states represent the collection in the way a statistical mixture of states is used in $R_{QT}\lfloor[\text{Physical Reality}]$. Thus for R_{KVR-QT}, $S_{\mathcal{L}(S_{corp})} = S_{\mathcal{L}(S_{doc})} \cup \ldots$, meaning that *there is a relationship in the representation* of R_{trad} elements 'document' and 'corpus' in R_{KVR-QT} as they conform to states and statistical mixture of states in R_{QT}. Conformance also applies within documents which can be represented as a convex collection of term states thus $S_{\mathcal{L}(S_{doc})} = S_{\mathcal{L}(S_{word})} \cup \ldots$, hence it can be said that $\mathcal{L}(S_{doc}) \approx_{conforms} \mathcal{L}(S_{corp})$ with respect to R_{KVR-QT} and therefore also with respect to R_{QT}. Thus, there is a statement in S_{corp} that has english translation 'language of R_{KVR-QT} can be used to model structures like me...', if this statement is true then it can be said that the language of S_{corp} 'conforms' to the conditions for representation in the framework R_{KVR-QT}.

In a QT framework required for modelling of structured document retrieval, one would require representation of structure in the Hilbert spce, for example by using subspaces of a space to represent the smaller elements of a document, to conform to the representation of

---

necessarily conform thus one requires 'tricks'; however, this it is still a crucial pursuit to deduce ways to represent arbitrary S_{match} in the QT Hilbert space formulation.

[12] Recall that a $\mathcal{L}$ are statements about the language of statements of a state, perception or function, and are embedded in states themselves, thus $S_{corp} = S_{\mathcal{L}(S_{corp})} \cup \ldots$ and maybe $S_{\mathcal{L}(S_{corp})} = [S_{\mathcal{L}(doc_i)}] \cup \ldots$.

a corpus thus $\mathcal{L}({}^{R_{QT}}S_{doc}) \approx_{conforms} \mathcal{L}({}^{R_{QT}}S_{corp})$, which means that operations traditionally done in the space by QT to both the corpus and document elements (as represented) would need to be consistent and correspond to a parallel meaning in R_{QT}. This is where the physics semantics are relevant, and with requiring conformance in terms of the representation languages and on the operational level for QT, a necessary implication is that there requires to be conformance at the semantic level. Recall that semantics are represented by english language statements and that the parameter (usually explicit) in each flow is a state itself holding among other information statements about the semantics of the flows. So one can define a set of statements about semantics for flows in structured document retrieval in $R_{thesis}\lfloor R_{trad}\lfloor [S_{struct.system} \cup \ldots]$, map their representation to vector or matrix operations in R_{QT} and then check if the semantics corresponding to these operations in R_{QT} match with the semantics of flows in $R_{thesis}\lfloor R_{trad}\lfloor [S_{struct.system} \cup \ldots]$[13].

6.3.0.5 Strategy for using R_{thesis}

If one examines the general model of structured documents (above), the relevant mathematical structures for representation, and adopts parts of the QT framework for each concept (in the structured document models), then it is easier to perform the following observation:

$$R_{QT}\lfloor R_{thesis}\lfloor R_{trad}\lfloor [S_{struct.system} \cup \ldots] \qquad (6.1)$$

than it is to do $R_{QT}\lfloor R_{trad}\lfloor [S_{struct.system} \cup \ldots]$. Thus, the research strategy for fully applying R_{thesis} is to start at the lowest level of state in the hierarchy, to represent inter-document relationships in terms of subspaces of $\mathcal{H}$, and then to check at each level if $\forall k \in [1 \ldots n], T_{QT_i}\big(\mathcal{L}([S_x] \subseteq^k S_{system})\big) \approx_{conforms} T_{QT_j}\big(\mathcal{L}([S_y] \subseteq^{k-1} S_x)\big)$ where $S_x \subset^k S_{system}$ denotes the kth level substates and 'n' is the total number of levels. If these all conform then one can say that the translations $[[T_{QT_i}], [T_{QT_j}]]$ are **confomorphic**, so R_{KVR-QT} is confomorphic for (static) document and terms states but does not suggest how to represent flows and thus is not confomorphic to R_{QT} for dynamics.

---

[13] Thus the concept of conformation refers to both representational conformance in the context of mathematics and to conformance of semantics in the context of physical reality in for R_{QT}.

6.4 Use of R_{QT} in light of R_{thesis}

6.4.1 Utility Measures, Convergence and Exhaustive Analysis

The term level, document level and general level utility measures are at a higher level than path based measures as they address application specific semantics of utility of documents and other objects/states, instead of working on the level of paths and flows. Thus, there are $|T|$ flows for each state ${}_{t}\psi$[14] as there are that many comparisons required to perform the utility functions[15], thus each $|[\gamma]_{\nu}|$(i.e utility measure over flows)$=|T|$. The utility is a flow between two states which correspond to sub-flows between component of these states. The energy type for term level utility is simply the label 'utility' or 'information need flow' and can be designated a value equal to the result of the utility function itself. For document level utility, one can assign the energy type to be 'potential difference' (see Section 4.3.4.7). The utility calculation is such that each flow that it consists of contains the value of intermediate calculations between sub-components of states such as documents or terms. Also $\nu : (a, b) :: C(a, b)$, i.e. all utility functions with two parameters are also convergence functions. Further, it is known that the *maximum convergence* for a collection $S_{corp} \subset^{k} S_{\Pi}$ is $S_{opt} = [S_{doc}]$. So given probabilities, one can do mixed dynamic and static analysis to see if the current IN is optimal for a collection, thus it can be said that there is *maximum shared knowledge* between a user and a system (or its corpus) when the user's IN as judged by *the method of assigning probabilities of relevance*, is best represented by a current state (see Chapter 4 for a representation). There is a symmetry in that the system also has IN and it has IN potential; meaning that there is certain information which it can represent well[16], and this is where its expertise lay[17]. Therefore for two future possible states, the utility $\nu({}_{1}\psi, {}_{2}\psi)$ employs two sets of probabilities[18] corresponding to the two states, and if the the second state has more utility then it is to be perceived as a more favourable future state than the first. Combining the interpretation of the values of the utility function $\nu({}_{1}\psi, {}_{2}\psi)$ in Chapter 4 with the interpretation of search in Chapter 3, if

---

[14] Recall that $|T|$ is the number of terms).

[15] This also refers to computation complexity upon translation to CS.

[16] Relative to other systems.

[17] This refers to the domain of the data.

[18] It can also be some other measure over states or documents which need not be a probability, see Chapter 4.

every process is a search then it can be said that the system (agent) knows more about that which pertains to the second state than to the first state.

6.4.2 Developing Quantitative Models in R_{QT}

In light of the previous section, it is established that there relationship between the concepts of convergence, human-human information retrieval (HHIR) and shared knowledge[19]. How does one vary the shared knowledge in search agents so that it includes non-corpus information on a modelling level? It is easier to perform this modelling with R_{thesis} (upon the elaboration of the framework of the prior chapter) than using pure QT as in Chapter 4. Consider that thus far there is the design ${}_tS_{\Pi} = {}_tS_{\psi} = [{}_tS_{doc_i}] = [[{}_tS_{t_j}]_i]$ and the required design is ${}_tS_{c\psi} = [{}_tS_{doc_i}] \cup [S_{context}]$. Firstly, one cannot assume that $\exists T, \mathcal{L}({}_tS_{c\psi}) \underset{T}{\rightarrow} \mathcal{L}(R_{QT})$, instead consider ${}_tS_{\Pi_c} = {}_tS_{\psi} \cup {}_tS_{context}$ which in R_{QT} can be represented as a superposition between two wave functions, this is a design through observation $R\lfloor {}_tS_{\psi} \lfloor [\text{user, system}, \ldots]$[20] where $R_{S_{\psi}}$ is an independent measuring device and thus the observable modelling technique of QT is being employed. If semantic measures are required, not only referring to counting of flows but measurements based on the content held by flow parameters, and further, the desire is to map to QT, then there must be conformance for the parameters of flows. Let S_{ϕ} denote the collective set of flow parameters, then it is required that for $S_{\phi} = [S_i]$, $S_i \underset{T}{\rightarrow} v \in \mathcal{H}$, i.e. that each flow parameter have a mapping onto an object on $\mathcal{H}$, and that they collectively conform to semantics[21]. The space or vector being mapped onto does not have to be the same basis as the basis documents are represented on, unless one wants to compare between documents and whatever S_i represents, so for comparisons it is required that $S_i \xrightarrow{\gamma} S_{doc_i}$ such that $|[\gamma]| = |[S_{doc_i}]|$. As there are measures over the state ${}_tS_{\psi}$ this implies that there are measures over $[S_i]$.

---

[19] Recall from the introduction about the concept of a social gap and search as inducing a sociological imagination. Convergence can be interpreted as sub-function of search, it is the quintessential function for sharing the knowledge (acquired through search) that leads to a sociological imagination. A social gap can be defined in relative terms with our framework by quantifying the convergence over some Πbetween two agents and showing that it is higher than between two other agents in another process thus indicating that the social gap is larger in the former process.

[20] So S_{ψ} holds all information about the search. In Chapter 4 the relevance state representation ψ also corresponds to this particular observation scheme as it modelled the entire search scenario in terms of system's view of user interests in terms of probabilities of relevance of words or documents.

[21] In general partial conformance is adequate, depending on the modelling that requires to be done.

6.4.3 Deducing Probabilities of Relevance

So far in the thesis it was assumed that the probabilities of relevance are deduced by classical means and then supplied to the Hilbert space representation for analysis, how would they be calculated in the same space? In order to illustrate how consider that an alternative observation scheme is required, instead of $R\lfloor_t S_\psi \lfloor[\text{user, system}, \ldots]$ one needs $R\lfloor[\text{user, system}, \ldots]$ with search state representation $S_\Pi = S_{sys} \cup \ldots$ then $S_{sys} = S_{doc} \cup S_{int}$ and $S_{int} = [S_{item}]$ (i.e. interface items). To do QT-like operations, one investigates how S_{item} changes, deduces the corresponding γ's, and then decides whether these changes are reversible or not, corresponding to whether one can use unitary operators or not. Assuming $U_{int} \subset S_{int}$ and $[U_{doc}] \subset S_{doc}$ are substates holding all information about the change of documents and interface items over the course of the search process then Pick($^jU_{doc_i}$) is a way for document i to change. One can describe paths such that given a particular interface change a path of system changes on documents occur $^jU_{int} \longrightarrow [^1U_{doc_3}, {}^5U_{doc_4} \ldots]$ which denotes that the change numbered 1 happens to document 3 and 5 to document 4. Thus one requires only to pick $^jU_{int}$ (perhaps randomly) and this would result in a path of corresponding changes (randomly or otherwise) chosen from a pre-defined set of possible changes (see Section 4.3.7 for designing this in R_{QT} in terms of unitary matrices). This is general way change occurs on a high-level and it corresponds to an application of change operators in R_{QT}.

In general, algorithms for making U's are $[\gamma]$ such that $U_{int} \xrightarrow{\gamma} U_{doc}$. A specific model would be one that defines U_{int} exactly, let $U_{int} = [U_{items}]$ and find translations such that $T_{Eng}(U_{items}) =$'interactive intentions'[Xie02] which correspond particular interactions that a user can do on the interface S_{int} and define a set of *orthogonal (in R_{QT}) intentions* such that each intention implies a change in a document. If $S_{doc_i} = [{}_iS_{static} \cup {}_iS_{dynamic} \cup \ldots]$ then can it be represented in QT? The wave-function technique of Section 4.3.4 is unable to deal with more than just the static and dynamic weights as it stands. The next section addresses the use of observables and traditional QT ways of modelling observables, this was referred to as the measurement method in Section 4.3.3.

6.4.4 Possible uses of the Measurement Method

If one knows all possible changes in the system, as represented by $S = [S_{poss_i}]$ then one needs to create a structure. Thus, if the context refers to time periods, click/interaction frequency, interaction count and information about other users (with whom collaboration has taken place for example), then for the first part of the context there are two operators $E_2 \perp E_1$ with respect to R_{QT} and $T_{eng}(\mathcal{L}(S_{E_2} \cup S_{E_2}))$='Before 6pm or After 6pm' to represent in binary value what time a particular interaction took place. For the second, $F_3 \perp F_2 \perp F_1$ with respect to R_{QT} is such that $T_{eng}(\mathcal{L}(S_{F_3} \cup S_{F_2} \cup S_{F_1}))$ ='clicks above 20, clicks=20 or clicks below 20' could represent the number of interactions. Finally for the third, $G_2 \perp G_1$ with respect to $R_{QT}, T_{eng}(\mathcal{L}(S_{G_2} \cup S_{G_1}))$='three other people have interacted with a particular document today or more than three people have interacted with it today'. Including these variations in context the state of search would entail the representation $S_\Pi = [[S_E], [S_F], [S_G]] \cup \ldots$. Each bit of information about a document is extractable (by the borrowed QT semantics) by defining appropriate operators that inscribe the pattern being sought in the document. In this way, the application of a matrix to another matrix representing the state of a system can be interpreted as *pattern matching*. For each design of the search state $\mathcal{L}_{QT+eng}(S_\Pi)$ = 'item$_i$: $[_jE_i], \ldots$ item$_n$: $[_jE_n]$'. Whether each E is compatible or not depends on the semantics, upon the establishment of compatibility one can represent them accordingly by vectors.

6.4.5 Formulating new search models

Traditional style search models can be devised either by considering the measures over states such as $\nu({}_1\psi, {}_2\psi)$ and ranking states or documents accordingly, or the measures over documents and states from Chapter 4 can be employed for ranking documents and states respectively, thus suggesting a family of models for retrieval. As can be seen from the following example, ranks change when both static and dynamic information is used to calculate the relevance score of a document according to the technique for combining these weights of the utility function $\nu(d_1, d_2)$.

Example 3. *[Ranking using Dynamic and Static weights] Let the probabilities of relevance*

for terms [dog, cat, rain, car] be [0.2,0.2,0.3,0.3]. And let the documents in static representation be $d_1 = 0.1|w_1\rangle\langle w_1| + 0.9|w_3\rangle\langle w_3|$
$d_2 = 0.8|w_1\rangle\langle w_1| + 0.2|w_2\rangle\langle w_2|$
$d_3 = 0.1|w_3\rangle\langle w_3| + 0.5|w_4\rangle\langle w_4| + 0.4|w_1\rangle\langle w_1|$
$d_4 = 0.9|w_2\rangle\langle w_2| + 0.1|w_4\rangle\langle w_4|$
Now using just dynamic information the score of documents would be
$Score(d_1) = (0.2 + 0.3)/2 = 0.25$
$Score(d_2) = (0.2 + 0.2)/2 = 0.2$
$Score(d_3) = (0.3 + 0.3 + 0.2)/3 = 0.267$
$Score(d_4) = (0.2 + 0.3)/2 = 0.25$
with rank $(d_3, [d_1, d_4], d_2)$*. However with dynamic static mix:*
$Score_d(d_1) = (0.2 * 0.1 + 0.9 * 0.3) = 0.29$
$Score_d(d_2) = (0.8 * 0.2 + 0.2 * 0.2) = 0.2$
$Score_d(d_3) = (0.1 * 0.3 + 0.5 * 0.3 + 0.4 * 0.2) = 0.26$
$Score_d(d_4) = (0.9 * 0.2 + 0.1 * 0.3) = 0.21$
with rank (d_1, d_3, d_4, d_2)*. Thus there is a difference in rank.*

The following method illustrates the way to deduce rankings for the example above and test models based on the utility measure and other measures over states/documents:

1. Obtain a probabilistic model (such as the binary voting model (see Section 2.2 for discussion)
2. Using (1), create probabilities of relevances $\mathbf{Pr}(Rel|d)$ for each document for a given time
3. Calculate the utility values for each document then rank
4. Compare ranks based on the previous relevance judgements. Rankings of (3) and the pure probabilistic method in (2) (without utility)[22]

This way of characterising dynamics corresponds to a flow of type $Pick$ as the flow is used to indicate selection of future states. It is also related to representation/design

---

[22] Note that the $\mathbf{Pr}(Rel)$ may already (semantically) include mixed static-dynamic related values, so there maybe *semantic overlap*. For example, the probabilistic model could include static (tf-idf) information but this does not change the usefulness of this method.

of the flows, i.e. as $e^{\jmath\theta}$ or $e^{\jmath[\theta_a+\theta_b]}$ which denotes the dynamic part of a state (that is separate from the static). By relations from Section 6.2, for example $w_{i,j}\left|w_j\right\rangle$ where $w_{i,j}\left|w_j\right\rangle = f([\gamma]) \subset S_{doc}$, this is statics as it is a measure of potential flows. Thus, it has been shown how one can make search models by considering the picking of states and by the way in which they are represented. For example, as complex numbers with static/dynamic combined representation or as arbitrary pieces of information $f(S_{[term,doc]})e^{\jmath\cdots}$ which is using R_{QT} and directly denotes the structure of S_{Π}. Note that it is easier to analyse all this in R_{thesis} than R_{QT} if only due to not having to think within the constraints of R_{QT} temporarily and instead in a type of 'unified modelling language' for R_{QT}, i.e. R_{thesis}. The thesis so far, has shown algorithms, new (simulation) models and corresponding metrics (such as those employing the utility measures). It must be noted that rank comparison is not seen as a significant accomplishment, as even if the subjectiveness of it is excused, one must acknowledge the limitation in such methods as one is not considering the dynamic change of information needs. There are ways to augment ranking methods to include factors pertaining to changing IN's (see [WJ04]), however, these ways to do not scale easily to account for arbitrary S_{Π} designs[23].

6.5 User Modelling

For user modelling, with respect to computer simulation of users, note that the notation presented in the prior chapter closely resembles computational/programmatic languages. It is assumed that the user's purpose is inscribed in a search goal[24], for example to want to see five specific documents. This goal can be represented as a function $G = G([S_{doc_i}])$ (a function of documents), or if it is not to have seen documents but perhaps to have seen some words on documents then this can be further abstracted by saying that a user's goal is to satisfy some statements which could itself depend upon the truth value of multiple functions (and corresponding statements). For example, HaveSeen$(D_1, D_3, D_5) \wedge$ HaveSeen('cat', 'dog', 'boat')$\wedge$HaveSeenInOrder(D_4, D_7, D_9). In order to address the

---

[23] As then using R_{trad} methods for modelling arbitrary scenarios can be understood as mapping detailed scenarios, which would be specified in R_{thesis} in terms of measures over sizes of potential paths and flow sets, onto traditional ranking models, thereby artificially 'over-fitting' them.

[24] Thus addressing the purpose in HHIR by representing it as the search goal.

issue of testing user models in search scenarios the general user type is as follows: A user sees the user interface therefore $S_{system} \longrightarrow_{\gamma_{view}} S_{int} \subset S_{user}$[25] then $S_{int} \longrightarrow_{\gamma} \ldots \longrightarrow S_{cog}$ where the cognitive state $S_{cog} = S_{memory} \cup S_{reasoning}$, then $S_{cog} \longrightarrow_{\gamma} \ldots \longrightarrow S_{int} \longrightarrow_{\gamma} S_{system}$.

Interfaces can be modelled by the interactive intentions in [Xie02] (mapped trivially to our notation) so that there are interface elements whose purpose or interactive intention is 'to locate' and other elements which are 'to check out' information, and others 'to see'. In R_{thesis}, these intentions are semantics for a particular flow type, this can be extended to model not only interactive intentions but deeper cognitive phenomena such as emotion and socio-psychological factors (see Fig. 6.4) . Thus, it can be said that a user wants to locate a document on 'cars' on the internet, then secondly to 'check out' the document linked to the 'cars' document; therefore there is a time element here. This time element is represented by a path with semantic labels instead of flow symbols, $P=$ (locate(cars), check out(linked)). For the purpose of shared knowledge or convergence the user needs to know something about the interactive intentions model in [Xie02] so it is assumed that the user is trained. This is comparable to going to a librarian and knowing what services they offer, for example, they can walk around the library to find a book, suggest information and suggest sources of information. Further, it is necessary to know what they will not do such as driving the user to another library to look there. Thus, it can be said that given such a user who has structure (checkout X, locate Y) with $X =$ [items for checking out], $|X| = n$ and $Y =$ [items to locate]$|Y| = m$, there are $n.m$ combinations and a goal i.e. 'search ends if $P =$ (check out,..,..)=true'. For each $(x \in X, y \in Y)$, one can measure the number of steps before the path P is tread (flow in P occur); one could also randomly select different (x, y) and take an average by using the evaluator. Randomness functions which are a type of flow, are important on two levels in R_{thesis}; first, to represent uncertainty in user models (and their subsequent simulations[26]) and secondly, for combinatorial analyses of search scenarios by considering average case behaviour of a simulation.

What does the evaluator look like? Recall from the ostensive chapter that the information

---

[25] which denotes the image of the interface reaching the user's senses.

[26] Given a set of possible future states randomness can be used to select a future state.

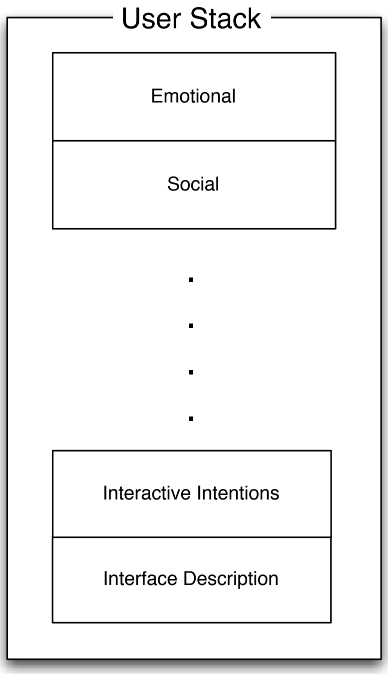


Figure 6.4. A stack design that considers high-level semantics for user modelling

need can in general be referred to as a need for experience, the user's search goal is therefore to gain an 'information experience'. In terms of traditional IR this means 'information experience' corresponds to document browsing. Each user model exhibits an information experience and this is represented by a state $S_{IE} \subset S_{cog}$. If some statement of S_{IE} is true then it is said that the experience has happened; a set of such statements being true would mean that the experience had taken place, while there maybe some uncertainty in this in the absolute sense it is not the case for a well-defined user model. Convergence of the schema of an experience which in R_{trad} would correspond to a set of relevant documents, and actual search experience is a sufficient condition to halt a simulation and check what changed in the process.

The idea of an information experience state comes from QT where one extracts something about a physical system by giving a prescription for measurement, this prescription is then applied to a state[27]. However, as in QT, one cannot directly observe the state of the system. All classical measures like energy, position and velocity are not directly available

---

[27] The prescription is like a query to nature (the system, by analogy) which returns the result of the measurement (it therefore *matches* the *measuring apparatus* to a physical reality) that corresponds to 'search results'; this analogy is accommodated by the broad definition of search given in Chapter 3.

to observation and instead are replaced by the 'all knowing' state vector[28]. The world of IR is also similarly hidden. Classical physics can be perceived a a special case of QT, as an approximation of reality. In IR, relevance or the 'click' that represents ones information need is similarly an approximation of the reality of the cognitive states of a user, or in terms of the concepts in this chapter, it can be said that flows and states at higher-levels are 'approximations' or generalisations, or the semantical interpretations, of those at lower levels.

What is S_{IE}? Either (1) it is a set of statements requiring to be true which can be complex and hierarchical as in Fig. 6.7; or (2) S_{IE} is a final state such that one keeps applying T_i to $S_{current}$ until $T_i(S_{current}) = S_{IE}$, the traditional vector space relevance feedback where a vector is rotated until it sufficiently represents user interests is a specific example of this transformation; or (3) S_{IE} could be a set of paths that require to have happened as would be observed by the evaluator agent meaning that the user agent has gone through certain changes. An *evaluation agent knows about user and system* so it is said that $K(S_{user}) \subset K(S_{eval})$ therefore convergence measurements are going held by and performed by S_{eval}[29] where R is doing the experiment. This assumes that the 'user is not conscious of themself' in the search, thus $K(S_{K(S_{user})}) \not\subset K(S_{user})$ in the views of the external observers which include the simulation software R_{sim}, the human investigator R_{exp} and R_{thesis}. The user therefore cannot perform the function of the evaluator, this is why the evaluator is conceptually justified as an individual entity, see Fig. 6.5. User's self-knowledge can be modelled as depicted by Fig. 6.6 where it is as if user had two roles, being of a self-evaluatory nature[30]. It is just an approach like (1) above where flows only occur upon conditions being met (see Fig. 6.7). Approach (2) is a way to discuss search goals that immediately gives the representational insight in translating it by $T_{\mathcal{L}(R_{vect})}(\mathcal{L}(R_{thesis}))$ thereby giving the impression of a 'vector moving through space' by rotation (or other transformation).

---

[28] Recall that the state of a system refers to all that is known about the system.

[29] Meaning that 'measuring for evaluation' denotes a flow type that denotes flows originating in the evaluation agent's general state (or a substate within it) as observed according to $R_{thesis}\lfloor R \lfloor [S_{eval} \cup \ldots]$.

[30] This model can also be used to model the notion of internal flux in [Cam00] where the user's internal state changes due to self-analysis, which is similar to needing to consider the change in measuring devices as they measure in R_{QT} which in turn relates back to this model since the evaluator component depicted in Fig. 6.6 is a measuring device in R_{thesis}.

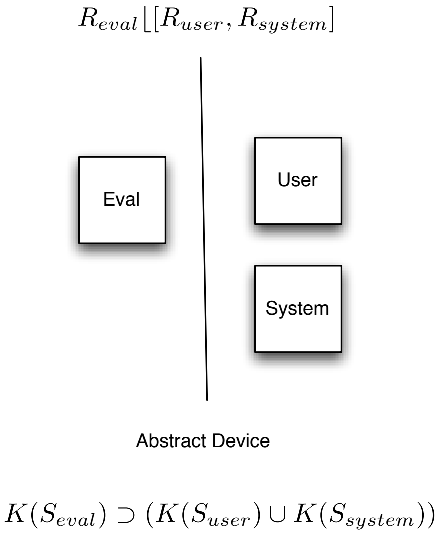


Figure 6.5. User's Self-knowledge denoted by a state observing itself through an evaluatory second role

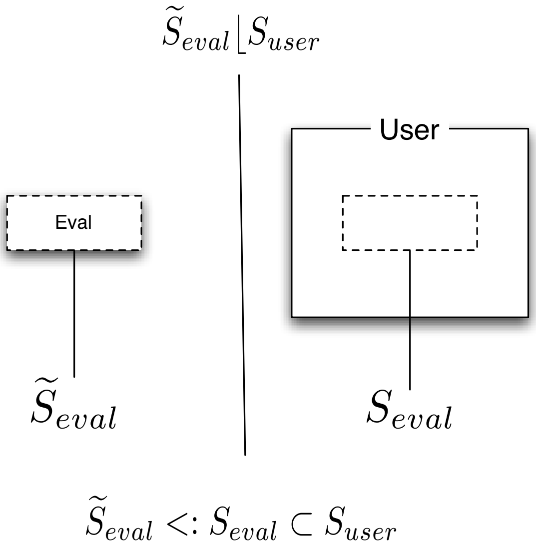


Figure 6.6. Evaluator as a separate agent that is an oracle relative to both simulated user and system

6.6 Evaluation Methods

In order to represent a user set and obtain relevance judgements consider that one user type is $P = (P_{int \to cog}, P_{cog \to int})$. More specifically let $P_{t_i} = S_{int} \longrightarrow S_{cog}$, going from interface to cognition but ignoring internal flows that detail changes in interface elements (thus it is a high-level view), and $P_{t_2} = S_{cog} \longrightarrow S_{int}$ but clicking 'non-optimal[31] icons'. Similarly define internal flows that represent deduction of if satisfaction of the need or S_{IE}

---

[31] Meaning that the interface elements which were not interacted with would have been more representative of the IN.

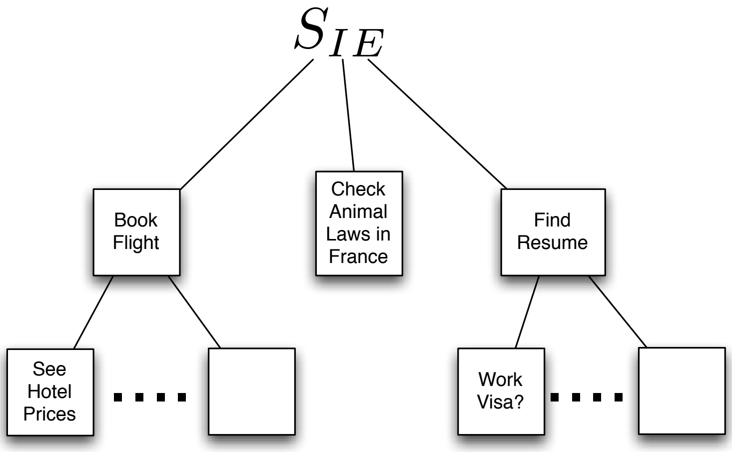


Figure 6.7. A possible substate structure in simple descriptive representation for the information experience that is sought S_{IE}

is adequately attempted[32], and a variation of topics $S_{IE} = [S_{cars}, S_{houses}, S_{...}, \ldots]$. Given such a setup, what measures alternative to the traditional precision/recall, can be devised to indicate the extent to which a search has met its goal? In general, measurements are associated with measuring states $S_M \subset^* S_{eval}$ (which are substates of the evaluator) and a general measuring state is $S_{M-gen} \subset S_M$ that has statements like "..'z' steps or $|P|$ or $f(|P|)$.." that allude to the structural or design properties of that being measured[33]. These are conceptually similar to quantum measuring devices except that they do not necessarily change on measurement or affect measurement (if that is not desired). Such a generalisation is an advantage of the abstraction being defined here over R_{QT} and therefore, is easier to work with, as there are lesser conditions. The state S_M and its substates contain highly linked statements, for example, 'how much' of 'what' (i.e. document) is 'what else' (i.e. 'read', 'relevant', 'interacted with'), as they relate objects with quantities and with other objects all of which are within the substates of the general search states; thus it requires to know precisely how the states being measured are structured. The English statements in S_M are akin to opinions that can be backed by mathematically write-able sub-parts such as *the system only understood the following user interaction at the level of*

---

[32] The state S_{IE} denoting information experience holds knowledge about if the user is satisfied or not thus $K(R_{satis}) \supset K(S_{IE})$ where R_{satis} is the perception containing a state that holds all traditional understandings of satisfaction of user need.

[33] This pertains to the number of flows, paths, path types, substates and substate types, and other conceptual categories by which a search is defined.

added to *..only 25 percent due to..* $f : (P, S_{sys}, S_{user} \ldots) \mapsto \mathbb{C}$[34]. Traditional evaluation is as follows, there are N users from whom relevance judgements are collected who are ideally experts or librarians who know the domain[35]. There is the system and results evaluator where the system is minimally represented as a set of documents and queries. There is also a result state $S_{user} \supset S_R = [q_i, [D_i]]$ for each user providing relevance judgements. The results are then fed to an evaluator in a new experiment, usually there is only minimal knowledge about $user_1, \ldots user_N$.

A new class of measures are generated by considering $f_m : P \mapsto \mathbb{R}$, specifically by considering subsets of paths. For example, the number of user interactions denoted by flows of type $\mathcal{FT}_{EXP}$ in the user compared to the number of flows between matching and corpus layers in the system, one can also formulate information theoretic measures to characterise the communication between user and system (see Section 5.4). The utility measure $\langle_{\text{current}}\psi|_{future_i}\psi\rangle$ for comparing current states to future states that are represented in a QT language is also a flow since $\psi <: S$ and a change from the current state to a potential future state is a flow, thus $\langle\phi_{\text{current}}|_{future_i}\psi\rangle <: f_m$ and *utility* $<: \mathcal{ET}$. As elaborated in Section 4.3.4, there are similarities between the semantics attributed to the utility function and the concept of potential difference which is a semantic in $\mathcal{L}\left(R_{class.phys}\right)$ according to the observation $R_{thesis}\lfloor R_{class.phys}$ and since there is a relation in $R_{classical.phys}$ between potential difference and energy, one can say that *utility* $<: \mathcal{ET}$ is an appropriate relationship that gives insight into what the utility measure is doing. There is also a relation between the utility measure and convergence it compares all document in one state with all documents in a potential future state, and since this comparison is given the semantic of 'similarity' as adopted from R_{trad}, it can be said that this is comparison of knowledge shared between current and possible future states thus is a function that should be explicitly[36] said to be of type convergence.

The utility measure can be used by the system model as a decision function and also by

---

[34] Usually the measures would output real numbers but if particular representations are used especially in R_{QT} then as shown in the previous chapter complex values can result.

[35] They could also be users taken from a population of expected users of the system depending on the aims of our experimental design

[36] Implicitly all measures on states or flows can be convergence functions as each state corresponds to a knowledge potential. Thus there is implicit comparison between the knowledges of these states; however, explicit convergence is the case where the knowledge being compared is known in totality. For example, in the utility function, it corresponds to finite sets of documents in each state with relevance values.

the user model. It is expected that for realistic user modelling the non-linear nature of human decision making, especially ‘random behaviour’, should be accommodated; due to its generality, complex decision functions can be represented and discussed in terms of flows and flow-semantics. For example, if Prospect Theory [KT79] is to be employed to model a simulated user’s decisions for picking future states then its utility function that refers to a set of prospects[37] and their value, considering behavioural heuristics and their probabilities, can be represented by the state $S_{\text{prospect decision}} = S_{\text{prospects}} \cup S_{\text{values}} \cup S_{\text{uncertainty}}$[38] and a flow (with corresponding flow type) that denotes the decision output. Each substate can be arbitrarily complex according to the detail of behavioural heuristics to be represented. The uncertainty and values for prospects can be calculated by comparing the information experience state to user’s understanding of the system’s state (SUP) by convergence measures. For example, if S_{IE} is represented as a set of words then the value of a prospective interaction with a system can be defined as a function of these set of words and words the user associates with interface items on ${}^{system}S_{int}$. Approximate users in this case can be characterised upon their peculiar prospect choosing behaviour as held in S_{values} and quantitatively classified by a function $f_m : (S_{values}, RU) \mapsto \mathbb{R}$ that checks how adequately S_{values} models real life behavioural heuristics[39] .

Combinatorial analysis of the number of possible paths in user and/or system models through analysing these f_m’s would aid in investigations of agent behaviour which would be especially useful from a user modelling perspective[40]. These measures would also help in performing exhaustive analysis of simulated scenario designs. Other classes of measures include $f_{m_2}(|p_i \subset P|)$, $f_{m_3}(|\gamma_x \in P|)$ that are measures on the lengths of paths and subpaths, and similarly there are measures to deduce the depth of a path design (to get an idea of design detail). Another other group of classes of measures are those that consider semantics, for example, if values for energy are given to flows like it is for flows/sinks on a discrete graph then there are measures $f_{m_4}(\text{Energy}(\gamma_x \in P))$. Thus there are naturally many variables to evaluate in a search scenario in R_{thesis} so that a scenario can be

---

[37] Note the overloading of the concept of the word ‘utility’, as the utility function of prospect theory is not the same as the one developed in this thesis.

[38] These substates are not necessarily mutually exclusive

[39] As S_{values} can also be represented through internal flows the measure of realism here would follow the design of evaluation measures specified at the start of this section

[40] As certain paths from the set of all paths in the user, denoting all the decisions they can make, may not be realistic, thus measures of realism or approximation error (see Section 3.7.3).

understood more deeply than with the usual evaluation metrics in R_{trad}.

In order to create of realistic search simulations there require to be experiments judging the level of realism of user models, given real users $[RU_1, RU_2, .., RU_n]$ one has to decide on $([(RU_i, UM_j)], \preceq)$ (a set with order) where (RU_i, UM_j) if UM_j is the closest match of RU_i. The human researchers of technical orientation are required to suggest $[UM_1, ..UM_n]$ based on their system models[41] and use them for investigations into search (i.e. without real users); then the user researchers' deduce which UM's are realistic by $R_{user.res}\lfloor[S_{RU_i} \cup S_{UM_j}]$[42]. If UM's are not found to be realistic enough then $[R_{user.res}, R_{tech}\lfloor_{R_{thesis}}]\lfloor[S_{RU_i} \cup S_{UM_j}]$ is necessary where the technical and user researchers require to discuss the reasons for ineffective models, other than this their interactions would be minimal due to the nature of simulation investigations. In this type of (non-classical) search investigation, one needs to create a search engine along with corresponding UM's and the UM's are to be represented so that they know what states exist within a search state in order to pick them as part of their information experience. Given that S_{sys} can be in the immediately next stage of one of $[S_a, S_b, S_c]$, the user model needs to know details about these possible future states unless it is a random user in which case it minimally needs to know the number of possible future states. However, each of these future states represent specifics about the search system; therefore user models need to know search engine specifics. What about the fact that real users do not know too many details about future states $[S_a, S_b, S_c]$? Real users may not initially know future states but they (can) learn. It is not suggested that they can learn everything about future states but that their knowledge about them improves/increases. This process of learning is of type convergence (see Section 5.4.8) which is something that is central to modelling dynamic behaviour in our abstraction. Thus realistic user models need to consider the learning process in real users to be good simulations.

---

[41] Recall from Section 5.4 the laws of association and separation.

[42] This experiment refers to the usual user-based experiments in R_{trad} and information seeking studies.

6.7 Structural Comparison of Models

The formalisation of R_{thesis} allows models represented in terms of states and flows to be compared on a structural level. For example, if the corpus of a vector space model is ${}^{\Pi}{}_{t}S_{vect.corp} = S_{ideal} \cup (\bigcup S_{D_i})$ where S_{ideal} corresponds to the learning vector that represents the current IN, then it can be said that at each $\tau > t$, the flows $\gamma_\tau <: \mathcal{FT}_{INT}$, due to user interaction, (traditionally) only cause S_{ideal} to change[43]. In a probabilistic model ${}^{\Pi}{}_{t}S_{prob.corp} = \bigcup S_{D_i}$, upon each user interaction, the probability of relevance of (potentially) all terms and documents change, thus each $\gamma_\tau <: \mathcal{FT}_{INT}$ causes $\xrightarrow{\gamma_\tau} S_{D_i}$, $\xrightarrow{\gamma_\tau} S_{t_{i,j}} \supset S_{D_i}$ and if term co-occurence information exists then it potentially causes $S_{1t_{i,j}} \xrightarrow{\gamma_\tau} S_{2t_{i,j}}$ i.e. the relevance of a term causes that of another term to change. Such a probabilistic model can be said to have more flows, and a more detailed relationship structure relative to the vector space model in which only S_{ideal} is modified per interaction. It also has potentially more types of flow, including flows between terms, documents and between groups of documents (i.e. in the case calculations such as $\mathbf{Pr}(Rel \text{ of } D_1 \vee D_2 | D_3, D_4, D_5)$ are required). Measures on the sets of different types of flows (as introduced in Section 5.4.4) are a way to characterise the relationships between data items in a model[44]. Statements that can be made from these characterisations can pertain to static relationships, such as the granularity of a state, i.e. the depth of the substate structure; or, it could refer to the dynamic relationships as indicated in the prior example, from which one may (simplistically) propose that the probabilistic model 'better interprets' the IN change at each time since 'more items change'. Whether such a statement is reasonable depends on a context specific analysis, that studies the particular need, e.g. through the corresponding S_{IE}, to deduce if changing all the data items per interaction is an effective way to interpret the corresponding IN or information experience. It may not be a good way if a particular interaction was not adequately representative of the user's IN but instead more of a careless interaction[45].

---

[43] Thus, the other states corresponding to document vectors do not change only the ideal vector changes. Note that this is only a simplistic characterisation of the traditional retrieval model that aims to illustrate how structural analysis can be done.

[44] For example, in the probabilistic case, due to the many flows and flow types it can be said that the model has more (traditional) information content than the vector space model.

[45] As it is reasonable to assume that not every interaction from a user is intended to be as representative of their IN as another.

Models can also be structurally compared by considering how they react to reversible flows, thus for each $\gamma_\tau <: \mathcal{FT}_{INT}$ that change the probabilities of terms and documents, or of the ideal vector, in the probabilistic and vector space models respectively, what does the inverse change $\overline{\gamma_\tau} <: \mathcal{FT}_{INT}$ do? Does $\overline{\gamma_\tau}$ undo the effect of γ_τ? For example, if a change in the relevance of a term causes the relevance of another term to change, i.e. $S_{cat} \xrightarrow{\gamma_\tau} S_{dog}$, then should it be that $S_{dog} \xrightarrow{\overline{\gamma_\tau}} S_{cat}$? That is, should the update in the probability of relevance of the term cat change the relevance of dog in the same way to how a change of relevance of term dog causes the relevance of cat to be updated[46]? The existence of reversible flows in a model indicate symmetric behaviour, and a quantification of this behaviour between terms could similarly be used to characterise the way a model depicts the relationship between data items. Reversible flows are less applicable in the structural comparison of traditional models which seldom make such relationships explicit [47] than they are in creating new models, for which it is proposed that they provide an important semantic for representing a crucial relationship between data items[48].

6.8 Discussion

It was proposed in the previous chapter that R_{thesis} depicts, on a general level, research in physics and computer science[49], but search *is* research and research *is* search, so if one is modelling search one is also modelling something about research except this is modelling automated or 'simulated' research[50]. This is summarised by Table 6.1 where one perception being inside another implies that the outer one is able to observe the inner in useful terms, for example, the observation $R_{trad}\lfloor R_{thesis}$ would be difficult as traditional IR does not

---

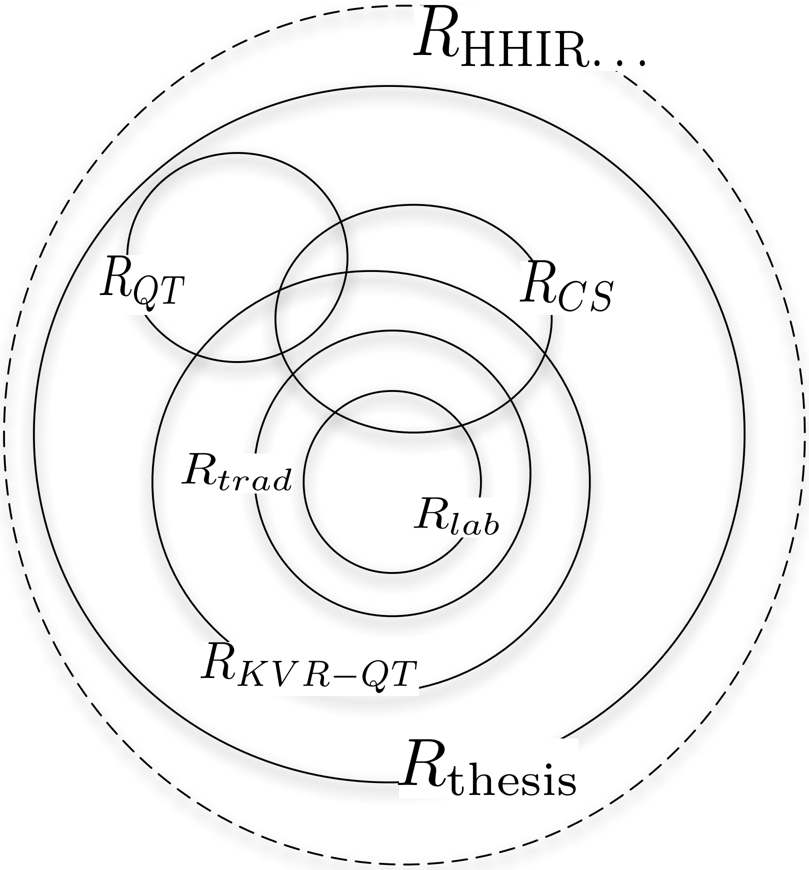
[46] This pertains to commutativity of change (see Section 4.3.7) applied to the (traditional) concept of updating term co-occurence information.

[47] Although methods of 'distributing' probabilities in the use inference networks [TC97] are a potential application of structural analysis.

[48] And also, between components and sub-components as illustrated in the previous chapter.

[49] This thesis did not explicitly explore $R_{trad} \mapsto R_{HHIR}$ but it is proposed that it is easier doing $R_{trad} \mapsto R_{thesis} \mapsto R_{HHIR}$ to find out how to better imitate HHIR. However, as studies of HHIR require the social sciences, the structured abstraction of a search process given by R_{thesis} is unlikely to complement $R_{soc.sci}$ insights as it stands, unless there is a good language for $R_{soc.sci}\lfloor[R_{thesis}\lfloor[\ldots]], [\ldots]]$ so it is clear how one can use our perception within $R_{soc.sci}$.

[50] This is since the process of modelling by $R\lfloor R_{thesis}\lfloor\ldots$ or otherwise is the same as *thinking about modelling* which is implicit simulation of a research and hence of a search scenario in the researcher's cognition.



$$R_{\text{Philosophical, Psychological, Social Scientific, Cognitive}\ldots} = R_{HHIR\ldots}$$

$$R_{HHIR\ldots} \widetilde{\lfloor} R_{\text{thesis}} \lfloor [R_{QT}, R_{KVR-QT} \lfloor R_{trad} \lfloor R_{lab}]$$

Figure 6.8. The perceptions explored in this chapter related with respect to observation operators $\lfloor$ or $\widetilde{\lfloor}$

contain the research vocabulary to address concepts in R_{thesis} whereas the converse has been shown in this thesis, also the frameworks corresponding to R_{HHIR} are defined as the most general observers but this is tentative at this time since it is unclear if the structured representation of R_{thesis} can be of much use for $R_{soc.sci}$ or other HHIR related perceptions, and until that is more clear the following (tentative) observation relations are proposed: $R_{thesis} \widetilde{\lfloor} R_{HHIR,\ldots}$[51] and $R_{HHIR,\ldots} \widetilde{\lfloor} R_{thesis}$[52]. In summary, to get to R_{QT} from R_{trad} it is said that if $S_1 \subset K(R_{KVR-QT})$ and $S_1 \subset K(R_{vect})$ and $S_1 \subset K(R_{QT})$ then $\exists s_a \in K(S_1) \cup K(S_{T_1})$ such that $T_1(s_a) = f(S_2)$ and $\exists s_b \subset K(S_1) \cup K(S_{T_2})$ such that

---

[51] Recall that R_{ost} addresses some general HHIR issues, and can observe both traditional IR and HHIR scenarios albeit from a high-level point of view.

[52] Clearly R_{thesis} is useful for some HHIR related phenomena related to cognition and such, however it is unclear if this applies in general and it is useful on the whole as it is for IR.

$T_2(s_b) = f(S_3)$, but note that $T_1(s_a) = f(S_2)$ implies that $T(s_A)$ is derivable from S_2 but 'deriving' corresponds to a research process. Therefore if one lets $f(S_2) = T(\dots)$ then this represents the idea that *'deriving' is search* and the function $f(S_2)$ implies that one has to *search for the map* f where the *search query is the state* S_f and the state $T(\dots) = search(f_2)$ *is the result of the search*[53]. The above pertains to the agents on the system side, what about on the user side/stack? What does it know? Due to symmetry, the answer is that it is the same, as one can perceive the user, and especially the simulated user, as a system; additionally, one can also see the user as a researcher.

$\mathbf{R_{thesis}}$ as a Pre-Formalisation R_{thesis} is a pre-formalisation, a notation set with usage instructions, a meta-model and discursive framework. In particular, it is a discursive framework that employs the QT way of analysing processes. It represents a researcher in the sense that the writer's mental/intuitionistic language (referring to the researcher deducing models) that cannot be fully expressed only 'projected'[54]. R_{thesis} is then the representation of a 'general researcher' who is able to 'by-pass' the self-referential issues at some abstract level[55] and is only possible due to its adoption of ideas from QT.

The notation presented in this thesis constitutes a language $\mathcal{L}(R_{thesis})$ to model (re)search and it allows one to specify the problem in the first place, thereby suggesting a solution to the generic definition problem. R_{thesis} offers resolutions on multiple levels for the conceptual problem by relating information need to energy, and flows to function applications, translations, and change, and the user problem to the problem of creating and applying approximate simulated users. The evaluation problem is still dependent on the user problem but R_{thesis} allows this dependence to be specified in a precise way, for example by using evaluatory measures on approximate user representations, and there are multiple ways through R_{thesis} to create evaluatory measures of differing types in the aim of measuring the information experience state S_{IE} which is a generalised form of the user satisfaction concept from R_{trad}.

---

[53] Thinking traditionally what is then the corpus in this case? It could be the set of all the known functions that can potentially denote the required translations.

[54] This is a subtle point saying that a researcher's understanding of the complex area of search is itself research knowledge that R_{thesis} is supposed to express except that this expression is approximate in the way a projection in QT is an approximation of the physical reality which is being projected into a perception.

[55] Perhaps by 'ignoring' it.

This chapter and Chapter 5 have drawn a similarity between state change, search, research, and knowledge ($K(\dots)$) change. Moreover, the knowledge of one state can be known by another, and thus it is said that one state is seen to know another according to a convergence function that is in the knowledge of an evaluator, and that the evaluator is also a researcher in this context. What is the use of all the notation introduced thus far, and how is it going to help us? In order to solve retrieval problems, one must be able to describe them relate them, be able to make new statements, and also define what it means to 'solve' them, this is precisely the definition problem. Also in order to use machines or people, one must be able to express problems in a language they will understand. The notation and ideas of R_{thesis} show how one gets (on a conceptual level and written form) from one idea to another, and provides some initial tools to represent IR research. Further, several concepts across these approaches were discussed relative to each other, a summary of these concepts with corresponding concepts in alternate perceptions are given in Table 6.1.

Representation in a particular form could allow some researcher in observation $R\lfloor\dots$ to derive new statements about a scenario whose deduction would otherwise be difficult and upon translation back to the source language can be said to 'solve' a problem or at least add new insight, thus representation of a problem could be crucial to finding a solution. As if one represents a problem in particular a form then:

1. A computer can be used to find new statements which may not be deducible otherwise. To use a conceptualisation similar to that of Chapter 3, a computer can be termed a *speed oracle* as it can be used to verify hypothesis and thus can *know* the truth of a hypothesis when a human using other methods does not yet know the truth of a statement.
2. A quantum computer can be used to find the truth of statements that are not tractable to find with a classical computer[56].
3. A human researcher with particular expertise can be used through associating statements by human intellect or intuition, that cannot be deduced by machines[57].

---

[56] Thus a quantum computer in terms of the concepts introduced in Chapter 3 would be a higher-level oracle such that $R_{\text{Quantum Computer}}\lfloor R_{\text{Classical Computer}}$.

[57] A human would be an oracle relative to the perspective of a computer accepting human interaction.

Recall from Chapter 2 that HHIR, specifically the HHIR in academic research, exhibits the optimal IR, thus it can be said that *research is the best search*; it is hoped that Chapter 5 and this chapter has alluded that R_{thesis} can model research. Modelling a group of researchers is itself a search scenario with multiple agents. It can denote researchers reading papers which involve in terms of R_{trad}, relevance feedback operations between the agents, or if looking at paper then if the paper has 'structure' or static knowledge then R_{thesis} provides notation to discuss how the researchers' knowledges converges towards the knowledge in the paper. This is similar to conventional search where the search engine or 'tool' takes over from 'hands', 'drawer', 'paper' and 'pen' that a researcher without a computer would use. In all these cases it is not talking about the live scenario but a simulated model of it, in fact any 'thinking of something' is a simulation of it[58]. All research involves simulation as all research involves thinking. R_{thesis} is a form in which research problems can be defined for research problems that can be seen to exhibit a process orientated flavour.

By addressing the general definition problem which is that of being able to define research problems in IR in an organised way and assuming simulation as the standard evaluation method, R_{thesis} thereby suggests a transformation of the user and evaluation problems into user research investigations. Further, the simple, typed set structure with flows of R_{thesis} offers new ways to understand IR in terms of physical semantics (of 'communication' and flows/changes) and through geometrical/mathematical semantics providing useful insights. Some examples of using R_{thesis} to create models were presented and their practical realisation is dependent on if one can translate that particular model $K(R_{thesis}) \supset K(R_{model})$ into R_{CS}. In general, R_{thesis} seems a good middle language between $R_{CS}, R_{QT}, R_{KVR-QT}, R_{\text{Object Programming Language}} \prec R_{CS}, R_{\text{State description approach}}$ and $R_{\text{Process description approach}}$. It is a teaching and research tool, a conceptualisation for search and schema for search scenario design.

Due to the generality of the framework proposed in this thesis, it accommodates modelling of arbitrary conceptualisations of a search system, structured data representation, and (hence) structured interaction and structured result representation[59].

---

[58] Therefore the deduction of R_{thesis} which induced thinking induced in turn a *simulation of research*.

[59] Structured interaction representation pertains to complex interactions, as in HHIR. It is further proposed that R_{thesis} supports modelling of complex filing strategies for S_{system} for aiding UUP and VIRP.

Perceptions			
R_{thesis}	R_{QT}	R_{CS}	R_{trad}
States, substates, statements, potentiality of concepts and documents, knowledge potential	Wave functions, convex mixture, entangled states, superpositions...	State ...	Various, Positions of vectors in R_{vect}, set of probabilities in $R_{prob} \ldots$
Flows, Paths	Operator	Functions	Feedback
Change, Information, Information Need, Energy	Evolution, Projection, Measurement, Energy...	State transition	Interaction, Document Select, result, query
Pick, Convergence, Expression, Interpretation, Observation	Heat, Types of Energy...	Program Compilation and Execution	Interactive Intentions, SUP, VIEP, UUP, QFP, Relevance
Systems, Evaluator, Simulated/Approximate Users, Shared Knowledge	Physical Systems	Models of computation, hardware models...	Real users, System
Pick new state, user-decision function, ostension, higher-ostension	Wave function collapse, super-selection...	State transition	Feedback, Interaction
Evaluation: Set-theoretic measures, Combinatorial, Information Theoretic	Projections, Density operators, inner-products...	Algorithmic complexity analysis	Precision/Recall based measures
Pseudocode, Technical English, ...	Technical English, Hilbert space theory, QT...	Programming languages and algorithms, Discrete Mathematics..	Technical English, pseudo code, ...
Oracles, Types of researcher....	Physical reality...	Halting problem, computability, self-referential issues	N/A

Table 6.1. Summary of similarities between concepts in R_{QT}, R_{CS}, R_{trad} and R_{thesis}

6.9 Summary

This chapter presented ways to model structured documents and the change of properties of structured documents in the course of a retrieval. It showed that the *information experience* concept of Chapter 3 can be represented in the states/flows notation and suggested techniques for generating user models in terms of information experiences. Several types of evaluation measures for the design of a search scenario in terms of states/flows were presented, some of which were related to the measures over states represented in QT formalism derived in Chapter 4. Methods for deriving probabilities of relevance, and hence creating traditional retrieval models were also briefly discussed. The discussion part of this chapter summarised the relationships between concepts from QT, CS, IR and the framework presented in this thesis, and showed the relationship between the presented framework and HHIR. It also re-addressed the practical, theoretical and conceptual benefits of translating search concepts between these frameworks which as hoped, is made easier by the formalisations presented in this thesis.

According to the framework introduced in this chapter, in order to use QT to model IR, one needs to deduce the translation:

$$T : ([[\text{Syntax}], [\text{Semantics}]] \subset K(R_{trad})) \mapsto \tag{6.2}$$
$$([[\text{Physics 'Syntax'}], [\text{Physics 'Semantics'}]] \subset K(R_{QT}))$$

Mapping the syntax here means that one can use the geometry, corresponding logics and measures in $\mathcal{H}$, but as discussed the physics semantics are also useful for modelling search. The main argument in this chapter and Chapter 5 was that it is easier doing $K(R_{trad}) \mapsto K(R_{thesis}) \mapsto K(R_{QT})$ than $K(R_{trad}) \mapsto K(R_{QT})$ or doing $K(R_{trad}) \mapsto K(R_{KVR-QT}) \mapsto K(R_{QT})$ [60]. Also we hope to have shown that it is easier doing $K(R_{trad}) \mapsto K(R_{thesis}) \mapsto K(R_{CS})$ than $K(R_{trad}) \mapsto K(R_{KVR-QT}) \mapsto K(R_{CS})$ or $K(R_{trad}) \mapsto K(R_{CS})$.

---

[60] Note that the discussion of Chapter 3, denoted by R_{ost} is precluded in R_{thesis}, thus the physical semantics used in discussing R_{ost} is also part of R_{thesis}.

> *—Search is the perennial yearning to be found ....I make no claims*
>
> Anonymous

7

Conclusions

7.1 Conclusions

In this section I list the contributions made by this thesis and outline the main conclusions. This thesis investigated the principles underlying the foundations of information retrieval. I outlined the four key problems in the area, the conceptual problem, definition problem, user problem and evaluation problem in Chapter 2. I then addressed each problem by analysing it from the perspective of the scientific framework underlying Quantum Theory. In the following paragraphs I list the contributions this work has made.

7.1.1 Ostensive search represents any search

This thesis argued that all information retrieval processes can be reduced to a generic ostensive retrieval process, and specifically the dynamics of a search process can be described as according to principle of higher ostension in Chapter 3. I introduced the concept of oracles and discussed the relative nature of observations about search, specifically distinguishing system and researcher points of view. I proposed that all processes are search processes and addressed the conceptual problem in this way. In particular, *search is not only a process but also the process by which all other processes are known, interpreted and expressed*. In IR, *search is both the object of study and the method of study*. The most external agents of a process always act ostensively relative to some observers viewpoint. An agent is relatively ostensive to an agent if the agent is unsure about the future interactions of the agent.

I argued that questions of idealness of a retrieval agent are not definable or addressable unless user's are simulated, suggesting the concept of approximate users. This work implies that it is inevitable that simulation be a normative evaluation strategy if extensive formal research is to be accommodated with respect to the IR research methodology presented in this thesis. I proposed that the user and systems have to be modelled to indicate that they share knowledge prior to a search, which suggests a symmetry in how a (simulated) user and a system ought to be modelled.

7.1.2 Techniques of Analysis from Quantum Theory

Using modelling techniques from QT, I argued that there are two main ways to model a search, first by its effects and the second by modelling a particular search phenomena such as relevance. These are not unrelated but separating them in this way provides insight that aid in the creation of models.

I developed models that characterise relevance by encoding static and dynamic search information. From this, I introduced measures of effectiveness (utility measures) that employ this encoding. I proposed that the QT formalism does not provide a way to predict how a search state will change, instead, one needs the equivalent of physical equations or laws to model[1] the dynamics of a search.

I showed that mathematical semantics such as commutativity that equate to conservation of a physical quantity, are useful for modelling search behaviour, specifically invariance in search process such as that pertaining to information need. I suggested that traditional IR is like classical physics whereas R_{thesis} accommodates a broader range of phenomena so corresponds more to quantum physics. A relation between physical semantics and IR semantics in the example of relating document-level utility measure to potential difference was given (see Section 4.3.4.7). A technique for using groups of unitary matrices to represent information need change was presented in Section 4.3.7 for designing simulations.

I showed that QT alone cannot address the definition problem of IR even though it provides a useful formal framework for modelling search phenomena.

7.1.3 General Representation Techniques

In Chapter 5, I propsed a notation and semi-formal representation technique incorporating QT semantics (from all QT postulates) for expressing search concepts that pertain to the general definition of search given in Chapter 3. The notation corresponds to a set-theoretic structure for specifying static details of a search, augmented with a graph structure for specifying the dynamics of a search.

---

[1] Thus one needs the IR equivalent of a Schrödinger equation and/or a field theory [Sak94] representing the semantics of the dynamics of a search.

Using this notation I defined the concept of shared-knowledge between agents, and introduced the concept of convergence by which a search can be generally characterised as a process by which knowledge of agents are modified. I gave several techniques for quantifying convergence between agents and between states of a search. In addition, I showed that the query formulation problem and system understanding problem can be specified as problems of convergence.

The presented representation offers insights from physics and computer science for modelling search. Therefore it allows concepts represented in its form to be translated into frameworks in these fields. Further, the representation accommodates complex models of cognition for modelling the user.

The representation language I proposed is an initial step to comprehensively tackling the definition problem of IR. Further, as it accommodates modelling of users (as discussed in Section 6.1), it is also an initial step to resolving the user and evaluation problems (as introduced in Section 2.6). The evaluation problem is additionally addressed by several types of measures presented in Chapter 5, Section 6.1, which suggest families of different measures (and corresponding semantics) to complement the traditional measures of effectiveness in IR.

From the perspective of traditional IR, I proposed techniques for formulating new search models for matching queries to documents using measures over documents, sets of documents, states of components or states of agents or searches over arbitrary time periods. These techniques make use of the utility measure developed in Chapter 4 where I suggested that this measure (which in general has several forms) is semantically related to measures of uncertain inference, and decision theoretic measures.

In Chapter 6, I suggested that user modelling relates to designing an interaction protocol for a system and that HHIR is the optimal interaction protocol. Further, I provided methods for modelling of structured document retrieval, analysis of requirements for translating a search scenario description to QT for formal analysis.

7.2 Future Work

As this thesis is a study into foundations, there are a significant number of possible future projects directly related to this work that are worthy of further investigation, or that stem as a consequence of the findings of this thesis. In the following a number of areas are outlined for possible future work.

A Philosophical Study The thesis proposed that search is intimately connected to research, and research to search. Search was also proposed to be the primitive tool for the individual computer user's development of a social perspective. Further, the higher ostensive principle proposed that all processes are search processes, and yet search is the process by which any process can be interpreted. Further work in the form of a comprehensive philosophical investigation of search, from a philosophy of science, philosophy of physics and philosophy of computer science perspective, would be necessary to definitively address the conceptual problem in IR. Philosophical development of the principle of higher ostension can be approached by considering its relation to methods of conceptualising participating observers in the QT context as suggested in [Sta07].

A Simulation Framework Given the various techniques for user, and system modelling, and corresponding evaluation measures, the next natural step would be to use these developed modelling tools to create a general (modelling and software) framework for simulating search. Techniques have been presented for analysing the design complexity of simulations, the realism of approximate users, and the complexity of a simulation, all of which are functions of measures on the sets of state changes in a search. These techniques can be employed to quantitatively classify simulation models.

Formal analysis of the Representation Language The semi-formal language presented is an intuitive one in which to represent search ideas, it is also a 'middle-language' between programmatic representation and a quantum theoretic state representation; it is necessarily ambiguous. A formal study of this language for the purposes of relating to formal computational languages, would be a useful investigation as it would directly address the practical side of the definition problem in IR.

Investigation of Utility Measures The measures developed in Chapter 4 are flexible in that they allow encoding of several types of details about a search, and the algebraic structure of $\mathcal{H}$ in QT allows these details to be combined is several useful ways. One avenue for further work is to investigate the formal relationships between these measures and decision theoretic measures such as the utility function from Prospect Theory [KT79], to find out if the utility function can be used to quantify a 'logical uncertainty principle' between states of search. A useful study would be to construct new models from the decision making techniques suggested by the utility measures, and to compare this according to the traditional retrieval evaluation strategies.

Appendices

A

Encoding

A.1 Probabilities into Angles

$$|{}_t\psi\rangle = \sum_{i \in \, {}_t\chi \, \subset \, V} \left({}^1p_{\,i} \prod_{\tau=1}^{t-1} {}_\tau\omega_i \right) |d_i\rangle \tag{A.1}$$

Which can also be written subsuming ${}^1p_{\,i}$ into a complex representation:

$$\begin{aligned}
|{}_t\psi\rangle = \sum_{i \in \, {}_t\chi \, \subset \, V} e^{\jmath\,({}_t\Theta_i)} \, |d_i\rangle \,, {}_t\Theta_i &= arg(\,{}^1p_{\,i}) + \sum_{\tau=1}^{t-1} {}_\tau\theta_i \\
\text{where we require that: } \sum_{i \in \, {}_t\chi \, \subset \, V} &\cos\left({}_t\Theta_i\right) = 1 \\
\Leftrightarrow {}_t\Theta_i = 2\pi({}_t\mathbf{Pr}(d_i)) \rightarrow \frac{1}{2\pi} \sum_{i \in \, {}_t\chi \, \subset \, V} &{}_t\Theta_i = 1 \\
\Leftrightarrow \sum_{i \in \, {}_t\chi \, \subset \, V} \mathbf{Pr}({}_td_i) = Re\left(\sum_{i \in \, {}_t\chi \, \subset \, V} e^{\jmath\,({}_t\Theta_i)} \right) &= 1
\end{aligned} \tag{A.2}$$

A.2 Document-Level Utility

In the following probabilities of relevance are mapped to a section of a sine curve that is approximately linear, the parameter (λ) is set according to the desired number of decimal places one requires to store a probability value thus it corresponds to the level of detail in the probabilities of relevance.

$$\text{Let: } \widehat{\mathbf{Pr}}\left({}_td_i\right) = \epsilon \mathbf{Pr}\left({}_td_i\right),\ \epsilon = 1 \times 10^{-\lambda} \iff \mathbf{Pr}\left({}_td_i\right) \approx_{\lambda} \frac{1}{\epsilon}\sin\left(\widehat{\mathbf{Pr}}\left({}_td_i\right)\right) \tag{A.3}$$

${}_t\Theta_i$ in Equation A.2 is set to $\widehat{\mathbf{Pr}}\left({}_td_i\right)$, by $\mathbf{Pr}\left({}_td_i\right)$ it is meant $\mathbf{Pr}\left(Rel|{}_td_i\right)$.

Given $\left|{}_td_1\right\rangle, \left|{}_td_2\right\rangle \in {}_t\chi$, $\left|{}_td_3\right\rangle, \left|{}_td_4\right\rangle \in {}_t\chi'$ where ${}_t\psi = \left(w_{1,1}\, e^{\jmath({}_t\Theta_1)} + w_{1,2}\, e^{\jmath({}_t\Theta_2)}\right)\left|w_1\right\rangle$, and ${}_t\widehat{\psi} = \left(w_{1,3}\, e^{\jmath({}_t\Theta_3)} + w_{1,4}\, e^{\jmath({}_t\Theta_4)}\right)\left|w_1\right\rangle$ where

$$\begin{aligned}\left\langle {}_t\psi \middle| {}_t\widehat{\psi} \right\rangle &= w_{1,1}\, w_{1,3}\, e^{\jmath({}_t\Theta_1 - {}_t\Theta_3)} + w_{1,1}\, w_{1,4}\, e^{\jmath({}_t\Theta_1 - {}_t\Theta_4)} \\ &+ w_{1,2}\, w_{1,3}\, e^{\jmath({}_t\Theta_2 - {}_t\Theta_3)} + w_{1,2}\, w_{1,4}\, e^{\jmath({}_t\Theta_2 - {}_t\Theta_4)}\end{aligned} \tag{A.4}$$

Taking one such summand one gets:

$$\Delta_{1,3,1} = w_{1,1}\, w_{1,3} e^{\jmath({}_t\Theta_1 - {}_t\Theta_3)} = \ w_{1,1}\, w_{3,1}\left[\cos\left({}_t\Theta_1 - {}_t\Theta_3\right) + \jmath\sin\left({}_t\Theta_1 - {}_t\Theta_3\right)\right] \tag{A.5}$$

The $\Delta_{i,j,k}$ in Equation A.5 for a fixed term $k = 1$ and in the case of two documents in each of ${}_t\chi'$ and ${}_t\chi$ is given inEquation A.5 which shows the results of the inner product between ${}_t\psi$ and ${}_t\widehat{\psi}$ where $w_{i,j}$ is the static weight of term i in document j. Generalising over all terms:

$$\Delta_{1,3,1} \approxeq w_{1,1}w_{3,1} + \jmath\left(\epsilon w_{1,1}w_{3,1}\right)\left(\mathbf{Pr}\left({}_td_1\right) - \mathbf{Pr}\left({}_td_3\right)\right) \tag{A.6}$$

which are traditional dot products between documents as in the vector space model for retrieval Equation A.7:

$$\left\langle {}_t\psi \middle| {}_t\widehat{\psi} \right\rangle \approxeq \sum_{i \in {}_t\chi, j \in {}_t\chi'} \left\langle {}_td_i | {}_td_j \right\rangle + \jmath\left(\epsilon\right) \sum_{i \in {}_t\chi, j \in {}_t\chi'} \left(\mathbf{Pr}\left({}_td_i\right) - \mathbf{Pr}\left({}_td_j\right)\right)\left\langle {}_td_i | {}_td_j \right\rangle \tag{A.7}$$

The taylor approximations are:

$$\begin{aligned} sin(x-y) &\approxeq x - y - \frac{1}{6}x^3 + \frac{1}{2}yx^2 - \frac{1}{2}y^2x + \frac{1}{6}y^3 \\ cos(x-y) &\approxeq 1 - \frac{1}{2}x^2 + yx - \frac{1}{2}y^2 \\ tan(x-y) &\approxeq sin(x-y) \end{aligned} \tag{A.8}$$

By Equation A.3 and taylor approximations we get:

$$\begin{aligned} Re\left(\Delta_{1,3,1}\right) &\approxeq w_{1,1}w_{3,1}\left[1 - \frac{\epsilon^2}{2}\left(\mathbf{Pr}\left({}_td_1\right) - \mathbf{Pr}\left({}_td_3\right)\right)^2\right] \\ Im\left(\Delta_{1,3,1}\right) &\approxeq (\epsilon)w_{1,1}w_{3,1}\left[\mathbf{Pr}\left({}_td_1\right) - \mathbf{Pr}\left({}_td_3\right)\right] \\ &\text{Ignoring relatively negligible, } \epsilon^2 \text{ terms we get:} \\ \Delta_{1,3,1} &\approxeq w_{1,1}w_{3,1} + \jmath\left(\epsilon w_{1,1}w_{3,1}\right)\left(\mathbf{Pr}\left({}_td_1\right) - \mathbf{Pr}\left({}_td_3\right)\right) \end{aligned} \tag{A.9}$$

The utility function captures this distinction Equation A.10 and presents itself as a measure for quantitatively comparing states based on dynamic and static search information:

$$\begin{aligned} \text{Let } \nu_{ij} &= \left(\mathbf{Pr}\left({}_td_i\right) - \mathbf{Pr}\left({}_td_j\right)\right)\left\langle {}_td_i \middle| {}_td_j \right\rangle \\ \text{Let } \nu_0 = \frac{1}{\epsilon} arg\left\langle {}_t\psi \middle| {}_t\widehat{\psi} \right\rangle &\approxeq \frac{\sum\limits_{i\in{}_t\chi, j\in{}_t\chi'} \nu_{ij}}{\sum\limits_{i\in{}_t\chi, j\in{}_t\chi'} \left\langle {}_td_i \middle| {}_td_j \right\rangle} \\ -1 \leq \nu_{ij} &\leq 1 \end{aligned} \tag{A.10}$$

$$\nu_0 : \begin{cases} = \infty & \sum\limits_{i\in{}_t\chi, j\in{}_t\chi'} \left\langle {}_td_i \middle| {}_td_j \right\rangle = 0 \\ \in [-1, 1] & \text{otherwise} \end{cases}$$

Taking account of special cases we say that utility is a generalisation of all three measures (which are further discussed in Section 4.3.4.6):

$$\textbf{Utility} = \begin{cases} \nu_0 : & \left\langle \psi \middle| \widehat{\psi} \right\rangle \text{ is a general case} \\ \nu_1 : & \left\langle \psi \middle| \widehat{\psi} \right\rangle \text{ is special case 1} \\ \nu_2 : & \left\langle \psi \middle| \widehat{\psi} \right\rangle \text{ is special case 2} \end{cases} \tag{A.11}$$

B

Representing terms as Signals

B.1 Introduction

A document can be interpreted as a *collection* of concepts, and can then also be represented as a *function* of concepts. Initially assume that every word in a *processed*[1] document is a *relatively reducible* concept. Define relatively reducible as meaning that there exists other items in the document 'about' the same concept, hence all these terms have the *potential* to be grouped relative to *either* all terms in a document *or* the corpus. Retrieval theory states that one can represent all documents of a corpus in a matrix where the $(i, j)^{\text{th}}$ entry is a binary membership value of the j^{th} term in the i^{th} document. Each row is then a document vector in a concept space spanned by all concepts in the corpus. Given the properties of synonymy and polysemy as a characteristic of the language in which the documents are represented, it makes more sense in grouping the dimensions of the space due to conceptual correlations between spanning vectors. Dimensional reduction algorithms generally work by identifying clusters differentiated on the initial high-dimensional space by a measure, then forming a set of reduced dimensions (assuming there clusters $\ll$ initial dimensions) from information derived from these clusters. This is the general idea behind spectral decomposition[2]/eigenvector decomposition/singular value-decomposition and is also a simple logical idea, in that, given a set of objects each possessing several

---

[1] Processed means: removed of stop-words.

[2] As introduced in Chapter 4.

different types of characteristics one can suggest the characteristic using which there will be an *acceptable*[3] partitioning of the items, the nature of acceptability being the variable. Latent semantic indexing [DCC+03] percieves the dimensions of the dimensionally reduced concept space of documents as pertaining to some hidden structure about the representation language, termed semantics. The idea correlates to signal processing in which any electric signal can be modelled, according to a calculus as consisting of *base wavelets*[4].

B.2 Encoding Terms

In signal analysis a wave has two basic properties, amplitude and frequency, that is, a signal is thought of as an entity that can be described by how often its observed and its strength on observation. At a first attempt one can say that in a corpus each term is a signal, with the frequency being a function of the number of documents in which it occurs and the amplitude being defined as a function of the intra and document frequency for each document. This separation of intra-document and inter-document frequencies correspond to the tf-idf weights based on the resolving power, specificity, exhaustivity and similar concepts in traditional retrieval. Representation of a term given some ordering on all terms and documents is detailed below.

First, given a document set D with $\|D\| = N$ and n_t as the number of documents in which term t occurs, the corpus frequency of a term is defined as:

$$f(t) = f\left(idf(t)\right) = \frac{1 + \log\left(\frac{N}{n_t}\right)}{1 + \log\left(N\right)}, \text{ setting } f(t) = 0 \text{ if } n_t = 0 \tag{B.1}$$

The value of the function has the following range:

$$\frac{1}{1 + \log\left(N\right)} \leq f(t) \leq 1 \tag{B.2}$$

---

[3] This refers to being conceptually acceptable corresponding to being a *quality dimension* in some conceptual space [GÖ0].

[4] Like a basis where each basis vector corresponds to a wave pattern. This also corresponds to harmonic analysis, which would perhaps be useful in studying linguistic patterns in text that are at a higher-level (semantically) than information theoretic conceptualisations that are based purely on frequency of occurrence.

Secondly in document D_p the q^{th} term is given a normalised intra-document frequency by any function:

$$tf(p,q) : \mathcal{N} \times \mathcal{N} \mapsto \{0,1\} \tag{B.3}$$

Finally the q^{th} term in document D_p is assigned a complex value:

$$T_{p\ ,q} \;=\; tf(p,q)\, e^{\jmath \phi_q}, \text{ where } \phi_q = 2\pi f(q) \tag{B.4}$$

One can replace the tf and idf (Equation B.1) measuring functions with alternative suitable functions as long as they are independent in that the tf measure should not be related to idf; keeping with the idea of the phase/amplitude of a wave being two separate properties. In Equation B.1, a rarer term gets a larger value for $f(t)$ which when substituted in Equation B.4 gives a larger angle ϕ. Intuitively, ϕ is the period of a term/signal generated by a term where the unit is a corpus, until a meaningful ordering can be concieved of the underlying collection, further parallels with signal theory is not possible as 'periodicity' would be an undefined concept.

A document can be represented as a sum/superposition[5] of its weighted term assignments or as a vector given bases ε_q:

$$D_p = \sum_q T_{p\ ,q} \text{ or } \sum_q \varepsilon_q T_{p\ ,q} \tag{B.5}$$

Defining a term/document matrix with its p^{th} column being the vector D_p, the similarity of documents need to be re-defined on such a space as the dot product between any two documents p_1, p_2 gives $\sum_q tf(p_1,q) * tf(p_2,q)$ thus excluding the idf. Before defining such a measure note that the complex conjugate of the idf part of $T_{p\ ,q}$, say a rare term, gives a value with a period for a common term according to a symmetry that is characteristic of the *complex roots of unity*. There exists many other such correspondences in such a representation. Given techniques, such as sampling and fourier analysis, it is proposed that more information can be gathered about the relations between terms and documents of a corpus.

Using dimensional reduction techniques one can represent documents in a complex semantic

---

[5] This refers to the superposition of signals as opposed to that of quantum states.

space, it would be interesting to see if there are any parallels between this and dimensional reduction techniques in signal analysis such as Fourier Decomposition[6]. Retrieval research has given documents a richer representation by considering relations between words, sentences and acknowledging document structure. The equivalents of these properties of a document in terms of signal representation would be denoted by a collection of functions[7] to define these richer properties in the representation of a document.

---

[6] An initial work on using dimensional reduction techniques from continuous spectra on data collections can be found in [Hoe03].

[7] Note that the notation of Chapter 5 accommodates document properties (as shown in Chapter 6) and can be used as a middle language for translating to a signal representation framework in the way it is used to translate to a quantum theory framework.

C

Notes on Evolution

C.1 Hamiltonians, Super-selection and Ostension

As introduced in Section 4.3.4, relevance feedback is accomplished by changing the state ψ upon observing feedback. Feedback is traditionally a query whose effect is logically deduced from information that includes $P(Q_i|\rho)$ (see Chapter 4). Any non-trivial operator can now be used to update the state upon feedback. Unitary operators U are one such class of operators with the useful property $U^* = U^{-1}$ allowing the undoing of a feedback and a diagonal form $U = \sum_j \beta_j |u_j\rangle\langle u_j|$ with eigenvalues lying in the unit circle, $\beta_j = e^{\jmath\theta} \in \mathbb{C}$. Inner products and hence lengths of vectors are invariant under unitary evolution. Thus the geometrical effect of applying U to some vector is a projection onto the eigenspaces of the operator followed by a rotation within the respective eigenspace.

The change of states over time due to relevance feedback is analogous to the dynamics of a physical system which can be given some structure. If there is some determinism in the evolution process then the set of unitary operators used in the evolution process form at least a one parameter semi-group $[{}_tU]$, if not a group. This is true if ${}_tU_TU = {}_{t+T}U$ (semi-group property) which means that the feedback at time t followed by a feedback at time T is equal to the feedback at a time $t + T$. Interpreting this more clearly requires that one define a parametric form for the unitary operators. According to [vF91] all such operators can be expressed ${}_tU = e^{-\jmath\mathbf{H}t}$ for some Hermitian operator $\mathbf{H}$. More generally, on solving the Schrödinger equation for a time dependent Hamiltonian, $\mathbf{H}$ is given a parameter [Gri02]: ${}_tU = e^{-\jmath\int_a^b \mathbf{H}(t)\,dt}$ where a and b denote the time at the start and end of the evolution period. It is required for $\mathbf{H}(t)$ to be properly defined so that the semi-group property is met. If one chooses the Hamiltonian so that it has time independent eigenspaces which are those spanned by the term bases of $\mathcal{H}$ then an useful form for the unitary operator is ${}_tU = \sum_j^{dim(\mathcal{H})} e^{-\jmath\int_a^b \beta_j(t)\,dt}|t_j\rangle\langle t_j|$ where the time dependence has moved to the eigenvalues.

Lets take the simple case of a document wave function represented as a superposition of the orthogonal term basis $|d_k\rangle = \sum_j^{dim(\mathcal{H})} w_{kj}\,|t_{ij}\rangle$. Then ${}_tU\,|d_k\rangle = \sum_j^{dim(\mathcal{H})} w_{kj}e^{-\jmath\delta_j(t,b,a)|t_{kj}\rangle}$ where $\delta_j(t,b,a) = \int_a^b \beta_j(t)\,dt$, which is simply a rotation of each co-ordinate of the document vector. Chapter 4 uses a parameter free unitary operator using a time-independent Hamiltonian with $\delta_j(t,b,a) = {}_t\beta_j$ where the time variable is used only as a label. In this case I just put in new values for ${}_t\beta_j$ each time I get relevance feedback. Instead of using a density operator for decision purposes I used a function of the inner product between two state wave functions $f(\langle {}_t\psi|{}_T\psi\rangle)$ or a utility function ν where ${}_t\psi$ is a state before feedback and ${}_t\psi$ after feedback. In this representation, the ${}_t\beta_j$ which hold probabilities of relevance for each term were coded in a specific way to get a reasonable output for the inner product between states. The rotations of the individual terms upon each feedback point are linearly proportional to the change in relevance of the respective term. The probability of relevance or its change is not deduced by the inner product but some other method (e.g. pure retrieval method).

In the density operator method, by contrast, the decision function is the trace which is used to generate a probability ($\mathbf{Pr}(M|\rho)$). The density operator is evolved by applying in the following way $\rho_b = U\rho_a U^{-1}$. The setting of the ${}_t\beta_j$ in the unitary evolution of the density operator needs to correspond with some previous judgement: $\mathbf{Pr}(M|\rho_a)$. For this to mimic the updating in some traditional relevance model one can run the traditional model to get the updated probability distribution for terms and/or documents, then create the new density ρ_b by solving equations $\mathrm{tr}(\rho_b\mathbf{Q}_i) = \langle\mathbf{Q}_i\rangle$ where $\langle\mathbf{Q}_i\rangle$ for single term queries is the probability of relevance of a term. However, in the general case when ρ_b may have complex entries, there are, for N terms N^2 entries [Blu81] to find, this means N^2 probabilities must be available for a complete representation.

The new density operator ρ_b will have the same diagonal entries as ρ_a if one assumes unitary evolution has taken place. If the system is in its initial state with a=1 then all entries in ρ_a are real, and if one assumes real entries for Q_i then only the real components of U's matrix entries will play a part in the calculations (the imaginary parts cancel). The unitary operator's components can be found by solving equations $\rho_b = U\rho_a U^{-1}$, from which the ${}_t\beta_j$ can be deduced. However, just finding U's to transform from two given density states is of no consequence, instead a particular relevance feedback method, such as the traditional Rocchio method [Roc71] or the method of the binary independence model, must be generalised *in terms of* a set of unitary operators with some parameters. The aim is to try to encapsulate any type of relevance feedback as a unitary operation, and if it fails then to investigate what it is about IR that makes it deviate from this particular character of QT evolution. There are other evolution types from QT to try, each with a rich set of properties and structure, some of which have deep physical interpretations. For IR purposes, one does not expect to have any structure on the space of unitary operators except in rare deterministic cases, thus it seems adequate to have parameter free evolution in the general case. However, then the idea of invariance/conservation of quantities (with respect to a group of unitary operators, see Section 4.3.7) as in [vF91] seems unclear without the context of parameters. Conservation of quantities is related to super-selection operators which seems like the tool for finely defining what dynamics are allowed or not, and it maybe of importance (but there is a difficulty in interpreting this in the sense of parameter free evolution). The external oracle of Chapter 3 can be said to 'super-select' states (ostensively) in that when the next state is not known in advance but a potentiality of states exist, the external observer breaks the superposition as per the concept of higher-ostension. This also corresponds to state collapse (see Chapter 4). In the case that the relative external observer, as the user is to the system, requires to be modelled as needs to be the case for search simulation then the semantic of super-selection or state-collapose is an useful association to the semantic of state-picking by an oracle as it indicates that an external influence was experienced, this was addressed in Chapter 3.

C.2 Using Histories

For modelling dynamics with the full force of QT one needs to consider all operations to measure things as observables thereby requiring that they correspond to operators with particular properties. Over time, state of a physical system changes as well as that of the observers of that state. State change can be modelled by the Born rule (see Chapter 4 and [Gri02]) for a few events but for > 3 events the concept of quantum histories are required, and these need to conform to conditions of consistency, thus particular search paths may not be consistent with others on a semantic level as inscribed in the properties of observation operators. In Chapter 5, IR is treated in a semi-formal manner such that it is easier to deduce what search paths there are in a search, their nature, the components in the path and the effect a search event has on them, semantic consistency is deduced through such analysis and is required knowledge if one is to map general search events to histories.

Quantum Histories can be used symbolically employing the random-walk model to quantitatively suggest the probability of possible future search paths. By using the QT formalism one could also model events logically so that a current document d_a to which is related document d_b and d_c can be denoted $d_a \longrightarrow d_b \vee d_c$. This approach allows a logic to be applied to the ostensive graph and would also be useful for general interface modelling and specifically to modelling search gestures. In terms of histories, the work in [Cam00] can be said to have employed two time histories, a particular state of affairs where a document selected is shown with other branching documents on the ostensive interface can be represented in terms of quantum histories as $Y_2 = [Y_1] \odot [d_1^{\text{next}}]$ [1] and $Y_3 = [Y_2] \odot [d_2^{\text{next}}]$. The representation suggests considering longer than two item histories, and perhaps not necessarily constant length histories such that complex past behaviours can be modelled.

---

[1] Where $\odot$ is a type of tensor product.

Bibliography

[AA94] D. Aerts and S. Aerts. Applications of quantum statistics in psychological studies of decision processes. *Foundations of Science*, 1:85–97, 1994. 36

[AG02] D. Aerts and L. Gabora. Contextualizing concepts using a mathematical generalization of the quantum formalism. *Journal of Experimental and Theoretical Artificial Intelligence*, 14:327–358, 2002. 36, 40, 116, 142, 150, 157, 175, 180, 188, 189, 190, 196

[AOC00] K. Sugihara A. Okabe, B. Boots and S. N. Chiu. *Spatial Tessellations - Concepts and Applications of Voronoi Diagrams*. John Wiley, 2000. 150

[AvR07] S. Arafat and C. J. van Rijsbergen. Quantum theory and the nature of search. In *Proceedings of the AAAI Symposium on Quantum Interaction*, pages 114–121, 2007. 12

[AvRJ05] S. Arafat, C. J van Rijsbergen, and J. Jose. Formalising evaluation in information retrieval. In *CoLIS Workshop on Evaluating User Studies in Information Access Fifth International Conference on Conceptions of Library & Information Science - Context: nature, impact and role.*, 2005. 3, 12, 15

[Axl99] S. Axler. *Linear Algebra Done Right*. Springer-Verlag New York Inc, 1999. 98, 101

[BC81] E. Beltrametti and G. Cassinelli. *The Logic of Quantum Mechanics*. van Nostrand, 1981. 104, 106, 141

[BCC$^+$98] N. J. Belkin, J. P. Carballo, C. Cool, S. Lin, S. Y. Park, S. Y Rieh, P. Savage, C. Sikora, H. Xie, and J. Allan. Rutgers trec-6 interactive track experience. In *Proceedings of the 6th Text REtrieval Conference (TREC-6)*, pages 597–610, Gaithersburg, Maryland, USA, 1998. 26

[Bla84] D. C. Blair. The data-document distinction in information retrieval. *Commun. ACM*, 27(4):369–374, 1984. 2

[Bla02] D. C. Blair. Knowledge management: Hype, hope, or help? *Journal of the American Society for Information Science and Technology*, 53(12):1019–1028, 2002. 2

[Blu81] K. Blum. *Density Matrix Theory and Applications*. Plenum Press, 1981. 250

[BOB82a] N. J. Belkin, R. N. Oddy, and H. M. Brooks. Ask for information retrieval: Part 1 background and theory. *Journal of Documentation*, 38(2), jun 1982. 36, 41, 150, 175, 190

[BOB82b] N. J. Belkin, R. N. Oddy, and H. M. Brooks. Ask for information retrieval: Part 2 results of a design study. *Journal of Documentation*, 38(3), sep 1982. 36, 113

[Buc92] M. K. Buckland. Emanuel goldberg, electronic document retrieval, and vannevar bush's memex. *Journal of the American Society for Information Science*, 43(4):284–294, 1992. 86

[Cam00] I. Campbell. *The Ostensive Model of Developing Information Needs*. Ph.D. dissertation, Department of Computer Science, University of Glasgow, 2000. 10, 27, 28, 29, 31, 35, 37, 45, 58, 64, 67, 115, 124, 140, 174, 179, 208, 220, 252

[Cas06] D. O. Case. *Looking for Information: A Survey of Research on Information Seeking, Needs and Behaviour*. Emerald Group, 2006. 2

[CBK96] C. Cool, N. J. Belkin, and J. Koenemann. On the potential utility of negative relevance feedback for interactive information retrieval. In *Proceedings of the 19th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland*, page 341, 1996. 26

[Che02] J. C. H. Chen. *Quantum Computation and Natural Language Processing*. Ph.D. dissertation, Department of Informatics, University of Hamburg, Germany, 2002. 40, 41

[CMZ05] W. B. Croft, A. Moffat, and J. Zobel. Similarity measures for tracking information flow. In *Proceedings of the ACM Fourteenth Conference on Information and Knowledge Management*, pages 517–524, 2005. 131

[Cor97] L. Corry. David hilbert and the axiomatization of physics (1894-1905). *Archive for History of Exact Sciences*, 51:83–198, 1997. 50

[CPB$^+$96] C. Cool, S. Park, N. J. Belkin, J. Koenemann, and K. B. Ng. Information seeking behaviour in a new searching environment. In *CoLIS2, Copenhagen*, pages 403–416, 1996. 21

[CRB98] M. Chalmers, K. Rodden, and D. Brodbeck. The order of things: Activity-centred information access. *Computer Networks and ISDN Systems*, 30(1-7):359–367, 1998. 29

[CvR96] I. Campbell and C. J. van Rijsbergen. The ostensive model of developing information needs. *CoLIS*, page 251268, 1996. 28, 29, 115

[DCC$^+$03] S. Dumais, E. Cutrell, J. J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff i've seen: a system for personal information retrieval and re-use. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 72–79, New York, NY, USA, 2003. ACM. 42, 126, 246

[DP02] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 2 edition, 2002. 168

[Fuh92] N. Fuhr. Probabilistic models in information retrieval. *The Computer Journal*, 35(3):243–255, 1992. 10

[G0̈0] P. Gärdenfors. *Conceptual Spaces*. MIT Press, 2000. 9, 150, 157, 246

[GBGL07] R. Geambasu, M. Balazinska, S. D. Gribble, and H.M. Levy. Homeviews: peer-to-peer middleware for personal data sharing applications. In *SIGMOD '07: Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 235–246, New York, NY, USA, 2007. ACM. 43, 44

[GHK$^+$03] G. Gierz, K. H. Hofmann, K. Keimel, J. D. Lawson, M. Mislove, and D. S. Scott. *Continuous Lattices and Domains*, volume 93 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2003. 168

[Gin00] H. Gintis. *Game theory evolving: a problem-centered introduction to modeling strategic behavior*. Princeton University Press, 2000. 85

[GJ79] M. Garey and D. S. Johnson. *Computers and Intractability*. Freeman, New York, 1979. 82

[Gri02] R. B. Griffiths. *Consistent Quantum Theory*. Cambridge University Press, 2002. 58, 97, 98, 104, 105, 106, 113, 141, 175, 209, 250, 252

[Gro96] L. K. Grover. A fast quantum mechanical algorithm for database search. In *STOC '96: Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219, New York, NY, USA, 1996. ACM. 54

[Har80] D. J. Harper. *Relevance Feedback in Document Retrieval Systems: An Evaluation of Probabilistic Strategies*. Ph.D. dissertation, Jesus College, Cambridge, England, 1980. 26

[Har92] D. Harman. Relevance feedback revisited. In *Proceedings of the 15th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Copenhagen, Denmark*, pages 1–10, 1992. 22

[Hoe03] E.C.M. Hoenkamp. Unitary operators on the document space. *Journal of the American Society for Information Science and Technology*, 54(4):314–320, 2003. 120, 248

[IJ05] P. Ingwersen and K. Jarvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, 2005. 2, 13, 20, 81, 154

[Ill71] I. Illich. *Deschooling Society*. Marion Boyars Publishers, 1971. 8

[Jau73] J. M. Jauch. *Are Quanta Real? A Galilean Dialogue*. Indiana University Press, 1973. 109

[Jon79] K. Spärck Jones. Search term relevance weighting given little relevance information. *Journal of Documentation*, 35(1):30–48, 1979. 26

[JSS00] B. J. Jansen, A. Spink, and T. Saracevic. Real life, real users and real needs: A study and analysis of users on the web. *Information Processing and Management*, 36(2):207–227, 2000. 22

[KT79] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(3):327–358, 1979. 129, 224, 239

[Mac69] D. M. Mackay. *Information, Mechanism and Meaning*. The M.I.T Press, 1969. 90, 91, 161

[Mar95] G. Marchionni. *Information Seeking in Electronic Environments*. Cambridge University Press, 1995. 2

[Mel07] M. Melucci. Exploring a mechanics for context aware information retrieval. In *Proceedings of the AAAI Symposium on Quantum Interaction (QI)*, 2007. 106

[Mil59] C. W. Mills. *The Sociological Imagination*. Oxford University Press, London, 1959. 3, 5

[Miz97] S. Mizzaro. Relevance: The whole history. *Journal of the American Society of Information Science*, 48(9):810–832, 1997. 20

[MW07a] M. Melucci and R. W. White. Discovering hidden contextual factors for implicit feedback. In *Proceedings of the 2nd Workshop on Context-based Information Retrieval (CIR)*, 2007. 106, 136

[MW07b] M. Melucci and R. W. White. Utilizing a geometry of context for enhanced implicit feedback. In *Proceedings of the Conference on Information and Knowledge Management (CIKM)*, 2007. 104, 106, 136

[NC00] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000. 58, 99, 107, 156, 175

[OS81] D. N. Osherson and E. E. Smith. On the adequacy of prototype theory as a theory of concepts. *International Journal of Cognitive Psychology*, 9:35–38, 1981. 36

[Pap94] C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, USA, 1994. 82

[Qui60] W. O. Quine. *Word and Object*. The MIT press, 1960. 28

[Qui69] W. O. Quine. *Ontological relativity and other essays*. Columbia University Press, 1969. 28

[RJ76] S. E. Robertson and K. Sprck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 41(4):288–297, 1976. 26

[RLvR01] I. Ruthven, M. Lalmas, and C. J. van Rijsbergen. Empirical investigations on query modifications using abductive explanations. In *Proceedings of the 24th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, USA*, pages 181–189, 2001. 26

[Rob90] S. E. Robertson. On term selection for query expansion. *Journal of Documentation*, 46:358–364, 1990. 26

[Rob03] S. Robertson. The unified model revisited. In *Workshop on Mathematical/Formal Methods in IR*, 2003. 31, 36, 46

[Roc71] J. J. Rocchio. Relevance feedback in information retrieval. In *The SMART retrieval system: experiments in automatic document processing*, page 313323. Prentice-Hall, US, 1971. 26, 251

[Rog87] H. Rogers. *The Theory of Recursive Functions and Effective Computability*. MIT Press, 1987. 62

[Ros99] K. H. Rosen. *Discrete Mathematics and Its Applications*. McGraw-Hill, 1999. 123, 156

[Sak94] J. J. Sakurai. *Modern Quantum Mechanics, Revised Edition*. Addison Wesley, 1994. 236

[San94] M. Sanderson. Word sense disambiguation and information retrieval. In *Proceedings of the 17th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland*, pages 142–157., 1994. 22

[SB90] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 44(1):288–297, 1990. 21, 25, 26

[Sta07] H. P. Stapp. *Mindful Universe: Quantum Mechanics and the Participating Observer*. Springer, 2007. 238

[SvR83] A. F. Smeaton and C. J. van Rijsbergen. The retrieval effects of query expansion on a feedback document retrieval system. *Journal of Documentation*, 35(1):30–48, 1983. 26

[TC97] H. Turtle and W. B. Croft. *Inference networks for document retrieval*, chapter Croft1997, pages 287–298. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997. 227

[tHPvdW96] A. H. M. ter Hofstede, H. A. Proper, and T. van der Weide. Query formulation as an information retrieval problem. *The Computer Journal*, 39(4):255274, 1996. 8, 22

[URJ03] J. Urban, C. J. Rijsbergen, and J. M. Jose. An adaptive approach towards content-based image retrieval. In *CBMI*, 2003. 8, 22, 25, 28, 29, 115

[VB07] E. M. Voorhees and L. P. Buckland, editors. *The Sixteenth Text REtrieval Conference Proceedings (TREC 2007).*, 500-274. National Institute of Standards and Technology, 2007. 48

[vF91] B. C. van Frassen. *Quantum Mechanics, An Empiricist View*. Clarendon Press, 1991. 249, 251

[vR96] C. J. van Rijsbergen. Quantum logic and information retrieval. In *2nd Workshop on Information Retrieval, Uncertainty and Logic*, IR Group, University of Glasgow, Scotland, 1996. 41

[vR00] C. J. van Rijsbergen. Another look at the logical uncertainty principle. *Information Retrieval*, 2(1):13–22, 2000. xviii, 36, 130

[vR04] C. J. van Rijsbergen. *The Geometry Of Information Retrieval*. Cambridge University Press, 2004. 11, 41, 52, 53, 55, 106, 109, 111, 141, 147, 152, 153, 154, 155, 157, 172, 202

[VV87] B. C. Vickery and A. Vickery. *Information Science in Theory and Practice*. Butterworths, 1987. 2

[Whe80] J. A. Wheeler. Pregeometry: Motivations and prospects. In A. R. Marlow, editor, *Quantum Theory and Gravitation*. Academic Press, 1980. 111

[Whi29] A. N. Whitehead. *Process and Reality*. Macmillan (New York), 1929. 7

[Wid03a] D. Widdows. A mathematical model for context and word-meaning. In *Fourth International and Interdisciplinary Conference on Modeling and Using Context, Stanford, California*, 2003. 40

[Wid03b] D. Widdows. Orthogonal negation in vector spaces for modeling word meanings and document retrieval. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan*, 2003. 40

[Wil97] J. Williams. Information science: definition and scope. In J. Williams and T. Carbo, editors, *Information Science: Still an Emerging Discipline*. Cathedral Publishing, 1997. 2

[Wit01] L. Wittgenstein. *Philosophical Investigations*. Blackwell Publishing, 1953/2001. 64

[WJ03] R. W. White and J. M. Jose. An approach for implicitly detecting information needs. In *CIKM, November 3-8, New Orleans, Louisiana, USA*, pages 504–507, 2003. 26

[WJ04] R. W. White and J. M. Jose. A study of topic similarity measures. In *SIGIR, Sheffield, South Yorikshire, UK*, 2004. 129, 131, 217

[WJvRR04] R. W. White, J. M. Jose, C. J. van Rijsbergen, and I. Ruthven. A simulated study of implicit feeback models. In *26th Annual European Conference on Information Retrieval, Sunderland, UK*, 2004. 54

[WP03] D. Widdows and S. Peters. Word vectors and quantum logic: Experiments with negation and disjunction. In *Appeared in Mathematics of Language 8, Bloomington, Indiana*, pages 141–154, June 2003. 40

[WRJ02] R. W. White, I. Ruthven, and J. M. Jose. The use of implicit evidence for relevance feedback in web retrieval. In *ECIR , 25-27 March, Glasgow, Scotland, UK*, 2002. 26, 45

[WRJvR05] R. W. White, I. Ruthven, J. M. Jose, and C. J. van Rijsbergen. Evaluating implicit feedback models using searcher simulations. *ACM Transactions on Information Systems (ACM TOIS)*, 23(3):325 –361, 2005. 25

[Xie02] H. Xie. Patterns between interactive intentions and information-seeking strategies. In *Information Processing & Management*, volume 38, pages 55–77, 2002. 53, 147, 182, 214, 218

[Zip49] G. K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, 1949. 114