Lecture Notes in Computer Science 3408

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David E. Losada Juan M. Fernández-Luna (Eds.)

Advances in Information Retrieval

27th European Conference on IR Research, ECIR 2005
Santiago de Compostela, Spain, March 21-23, 2005
Proceedings

Springer

Volume Editors

David E. Losada
Universidad de Santiago de Compostela
Departamento de Electrónica y Computación
Campus sur s/n, Universidad de Santiago de Compostela
15782 Santiago de Compostela, Spain
E-mail: dlosada@usc.es

Juan M. Fernández-Luna
Universidad de Granada
E.T.S.I. Informática
Departamento de Ciencias de la Computación e Inteligencia Artificial
C/Periodista Daniel Saucedo Aranda, s/n, 18071 Granada, Spain
E-mail: jmfluna@decsai.ugr.es

Library of Congress Control Number: 2005921726

CR Subject Classification (1998): H.3, H.2, I.2.3, I.2.6, H.4, H.5.4, I.7

ISSN 0302-9743
ISBN 3-540-25295-9 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11406761 06/3142 5 4 3 2 1 0

Preface

Welcome to Santiago de Compostela! We are pleased to host the 27th Annual European Conference on Information Retrieval Research (ECIR 2005) on its first visit to Spain.

These proceedings contain the refereed full papers and poster abstracts presented at ECIR 2005. This conference was initially established by the Information Retrieval Specialist Group of the British Computer Society (BCS-IRSG) under the name "Annual Colloquium on Information Retrieval Research." The colloquium was held in the United Kingdom each year until 1998, when the event was organized in Grenoble, France. Since then the conference venue has alternated between the United Kingdom and Continental Europe, reflecting the growing European orientation of ECIR. For the same reason, in 2001 the event was renamed "European Conference on Information Retrieval Research." In recent years, ECIR has continued to grow and has become the major European forum for the discussion of research in the field of information retrieval.

ECIR 2005 was held at the Technical School of Engineering of the University of Santiago de Compostela, Spain. In terms of submissions, ECIR 2005 was a record-breaking success, since 124 full papers were submitted in response to the call for papers. This was a sharp increase from the 101 submissions received for ECIR 2003, which was the most successful ECIR in terms of submissions. ECIR 2005 established also a call for posters, and 41 posters where submitted. Paper and poster submissions were received from across Europe and further afield, including North America, South America, Asia and Australia, which is a clear indication of the growing popularity and reputation of the conference. All papers and posters were reviewed by at least three reviewers. Out of the 124 submitted papers, 34 (27%) were accepted; 17 (41%) posters were accepted.

Students are well represented, since 22 out of 34 full papers and 10 out of 17 posters involve a full-time student as the primary author, which means that the traditional student focus of the conference has been very well preserved.

The increasing presence of research papers from leading companies it is also remarkable.

We had an outstanding set of research contributions this year, reflecting the full range of information retrieval research areas. The proceedings start with two invited papers, from Keith van Rijsbergen and Ricardo Baeza-Yates. van Rijsbergen's work shows how logic emerges from the geometry of the popular vector space model. Baeza-Yates proposes two applications of analyzing and clustering queries stored in server logs of search engines and website logs. The topics covered by the papers and posters include peer-to-peer systems, formal models, text summarization, classification, fusion, user studies, evaluation, efficiency issues, image and video retrieval, web IR, and XML retrieval.

The success of ECIR owes a lot to many individuals involved in the reviewing tasks. We are deeply grateful to all involved in this process for their dedication and professionalism in meeting the very tight deadlines. We would like to extend our warm thanks to the researchers who submitted their results for consideration. Many thanks also to our keynote speakers Keith van Rijsbergen and Ricardo Baeza-Yates for agreeing to present at ECIR 2005. We are also extremely grateful to Fabio Crestani, Pia Borlund and Gianni Amati for facing the difficult task of deciding which student paper deserved the Best Student Paper Award. A special word of thanks is extended to Fabio Crestani, who has supported us from the very beginning of the bidding process.

We wish also to thank the companies and institutions who sponsored ECIR 2005: the Information Retrieval Specialist Group of the British Computer Society (BCS-IRSG), the University of Granada, the Council of European Professional Informatics Societies (CEPIS), Microsoft Research, Sharp Laboratories of Europe, Ltd., and the European Research Consortium for Informatics and Mathematics (ERCIM).

We would also like to thank the members of the Local Organizing Committee for their hard work over many months. They enthusiastically supported us in every small task related to the conference. Not all these persons may be visible to conference participants but the efforts of all are invaluable in making the conference a success.

Most of all, we would like to thank our wives, Maria and Nuria, for their endless patience and tolerance through the long hours dedicated to ECIR.

January 2005

David E. Losada
Juan M. Fernández-Luna

Organization

ECIR 2005 was organized by the University of Santiago de Compostela, with the collaboration of the University of Granada, and under the auspices of the Information Retrieval Specialist Group of the British Computer Society (BCS-IRSG).

Local Organizing Committee

Abraham Otero, University of Santiago de Compostela
Adolfo Riera, University of Santiago de Compostela
Alberto Bugarín, University of Santiago de Compostela
David López Moreno, University of Santiago de Compostela
Félix Díaz Hermida, University of Santiago de Compostela
José Luis Correa, University of Santiago de Compostela
Juan Carlos Vidal, University of Santiago de Compostela
Manuel Mucientes, University of Santiago de Compostela
María Pilar G. Souto, University of Santiago de Compostela
Paulo Félix, University of Santiago de Compostela
Purificación Cariñena, University of Santiago de Compostela
Roberto Iglesias, University of Santiago de Compostela

Programme Committee

David E. Losada, University of Santiago de Compostela, Spain (Chair)
Juan M. Fernández-Luna, University of Granada, Spain (Chair)

Andrew MacFarlane, City University London, United Kingdom
Alan Smeaton, Dublin City University, Ireland
Alvaro Barreiro, University of A Coruña, Spain
Anastasios Tombros, Queen Mary University of London, United Kingdom
Andreas Rauber, Vienna University of Technology, Austria
Arjen de Vries, CWI, Netherlands
Ayse Göker, Robert Gordon University, United Kingdom
Barry Smyth, University College Dublin, Ireland
Claudio Carpineto, Fondazione Ugo Bordoni, Italy
Djoerd Hiemstra, University of Twente, Netherlands
Dunja Mladenić, Jožef Stefan Institute, Slovenia
Eero Sormunen, University of Tampere, Finland
Enrique Herrera-Viedma, University of Granada, Spain
Fabio Crestani, University of Strathclyde, United Kingdom
Fabrizio Sebastiani, National Council of Research, Italy

Gabriella Pasi, National Council of Research, Italy
Gareth Jones, Dublin City University, Ireland
Giambattista Amati, Fondazione Ugo Bordoni, Italy
Gloria Bordogna, National Council of Research, Italy
Henrik Nottelmann, University of Duisburg-Essen, Germany
Hugo Zaragoza, Microsoft Research, United Kingdom
Iadh Ounis, University of Glasgow, United Kingdom
Ian Ruthven, University of Strathclyde, United Kingdom
Jane Reid, Queen Mary University of London, United Kingdom
Jesús Vegas, University of Valladolid, Spain
Joe Carthy, University College Dublin, Ireland
Joemon Jose, University of Glasgow, United Kingdom
Josiane Mothe, Institut de Recherche en Informatique de Toulouse, France
Julio Gonzalo, UNED, Spain
Jussi Karlgren, Swedish Institute of Computer Science, Sweden
Kalervo Järvelin, University of Tampere, Finland
Kees Koster, Radboud University Nijmegen, Netherlands
Keith van Rijsbergen, University of Glasgow, United Kingdom
Leif Azzopardi, University of Glasgow, United Kingdom
Luis de Campos, University of Granada, Spain
Marcello Federico, ITC-irst, Italy
Margaret Graham, Northumbria University, United Kingdom
Mark Girolami, University of Glasgow, United Kingdom
Mark Sanderson, University of Sheffield, United Kingdom
Massimo Melucci, University of Padova, Italy
Matthew Chalmers, University of Glasgow, United Kingdom
Michael Oakes, University of Sunderland, United Kingdom
Micheline Beaulieu, University of Sheffield, United Kingdom
Mohand Boughanem, Université Paul Sabatier, France
Monica Landoni, University of Strathclyde, United Kingdom
Mounia Lalmas, Queen Mary University of London, United Kingdom
Norbert Fuhr, University of Duisburg-Essen, Germany
Peter Ingwersen, Royal School of Library and Information Science, Denmark
Pia Borlund, Royal School of Library and Information Science, Denmark
Ricardo Baeza-Yates, ICREA-Universitat Pompeu Fabra, Spain, and
University of Chile, Chile
Sándor Dominich, University of Veszprém, Hungary
Sharon McDonald, University of Sunderland, United Kingdom
Stavros Christodoulakis, Technical University of Crete, Greece
Thomas Roelleke, Queen Mary University of London, United Kingdom
Tony Rose, Cancer Research UK, United Kingdom
Ulrich Thiel, Fraunhofer IPSI, Germany
Umberto Straccia, National Council of Research, Italy
Victor Poznanski, Sharp Laboratories of Europe, United Kingdom
Wessel Kraaij, TNO TPD, Netherlands

Best Student Paper Award Committee

Fabio Crestani, University of Strathclyde, United Kingdom (Chair)
Giambattista Amati, Fondazione Ugo Bordoni, Italy
Pia Borlund, Royal School of Library and Information Science, Denmark

Additional Reviewers

Alessandro Moschitti, University of Rome "Tor Vergata", Italy
Alex Bailey, Canon Research Centre Europe Ltd., United Kingdom
Amanda Spink, University of Pittsburgh, USA
Andreas Pesenhofer, Electronic Commerce Competence Center, Austria
Andrew Trotman, University of Otago, New Zealand
Antonio Ferrández, University of Alicante, Spain
Basilio Sierra, University of the Basque Country, Spain
Ben He, University of Glasgow, United Kingdom
Benjamin Piwowarski, University of Chile, Chile
Birger Larsen, Royal School of Library and Information Science, Denmark
Börkur Sigurbjörnsson, University of Amsterdam, Netherlands
Cathal Gurrin, Dublin City University, Ireland
Chris Stokoe, University of Sunderland, United Kingdom
Christof Monz, University of Maryland, USA
David Hull, Clairvoyance Corporation, USA
Dawei Song, Distributed Systems Technology Centre, Australia
Donald Metzler, University of Massachusetts, USA
Edda Leopold, Fraunhofer Gesellschaft, Germany
Edie Rasmussen, University of British Columbia, Canada
Fabio M. Zanzotto, University of Milano-Bicocca, Italy
Fernando Diaz, Center for Intelligent Information Retrieval,
University of Massachusetts, USA
Fernando Llopis, University of Alicante, Spain
Fidel Cacheda, University of A Coruña, Spain
Filippo Portera, University of Padova, Italy
Franciska de Jong, University of Twente, Netherlands
Gabriella Kazai, Queen Mary University of London, United Kingdom
Giorgio M. Di Nunzio, University of Padova, Italy
Hyowon Lee, Dublin City University, Ireland
Jamie Callan, Carnegie Mellon University, USA
Jean-Pierre Chevallet, CLIPS-IMAG, France
Jesper Schneider, Royal School of Library and Information Science, Denmark
Jian-Yun Nie, University of Montreal, Canada
José M. Gómez Hidalgo, Universidad Europea de Madrid, Spain
Juan Huete, University of Granada, Spain
Leonardo Candela, National Council of Research, Italy
Luo Si, Carnegie Mellon University, USA

Manuel Lama Penín, University of Santiago de Compostela, Spain
Miguel A. Alonso, University of A Coruña, Spain
Nicola Orio, University of Padova, Italy
Oscar Cordón, University of Granada, Spain
Pablo de la Fuente, University of Valladolid, Spain
Pasquale Savino, National Council of Research, Italy
Patrick Gallinari, Université Pierre et Marie Curie, France
Patrick Ruch, University Hospitals of Geneva, Switzerland
Pavel Calado, INESC-ID, Portugal
Pertti Vakkari, University of Tampere, Finland
Rafael Berlanga, Universitat Jaume I, Spain
Roberto Basili, University of Rome "Tor Vergata", Italy
Roi Blanco, University of A Coruña, Spain
Ross Wilkinson, CSIRO, Australia
Ryen White, University of Maryland, USA
Theodora Tsikrika, Queen Mary University of London, United Kingdom
Toni Rath, Center for Intelligent Information Retrieval, University of Massachusetts, USA
Vanessa Murdock, Center for Intelligent Information Retrieval, University of Massachusetts, USA
Vassilis Plachouras, University of Glasgow, United Kingdom
Xiaoyong Liu, Center for Intelligent Information Retrieval, University of Massachusetts, USA

Sponsoring Institutions

Table of Contents

Keynote Papers

A Probabilistic Logic for Information Retrieval
C.J. 'Keith' van Rijsbergen 1

Applications of Web Query Mining
Ricardo Baeza-Yates 7

Peer-to-Peer

BuddyNet: History-Based P2P Search
Yilei Shao, Randolph Wang 23

A Suite of Testbeds for the Realistic Evaluation of Peer-to-Peer Information Retrieval Systems
Iraklis A. Klampanos, Victor Poznański, Joemon M. Jose, Peter Dickman 38

Federated Search of Text-Based Digital Libraries in Hierarchical Peer-to-Peer Networks
Jie Lu, Jamie Callan 52

Information Retrieval Models (I)

'Beauty' of the World Wide Web—Cause, Goal, or Principle
Sándor Dominich, Júlia Góth, Mária Horváth, Tamás Kiezer 67

sPLMap: A Probabilistic Approach to Schema Matching
Henrik Nottelmann, Umberto Straccia 81

Encoding XML in Vector Spaces
Vinay Kakade, Prabhakar Raghavan 96

Text Summarization

Features Combination for Extracting Gene Functions from MEDLINE
Patrick Ruch, Laura Perret, Jacques Savoy 112

Filtering for Profile-Biased Multi-document Summarization
Sana Leila Châar, Olivier Ferret, Christian Fluhr 127

Automatic Text Summarization Based on Word-Clusters and Ranking Algorithms
Massih R. Amini, Nicolas Usunier, Patrick Gallinari 142

Comparing Topiary-Style Approaches to Headline Generation
Ruichao Wang, Nicola Stokes, William P. Doran, Eamonn Newman, Joe Carthy, John Dunnion 157

Information Retrieval Methods (I)

Improving Retrieval Effectiveness by Using Key Terms in Top Retrieved Documents
Lingpeng Yang, Donghong Ji, Guodong Zhou, Yu Nie 169

Evaluating Relevance Feedback Algorithms for Searching on Small Displays
Vishwa Vinay, Ingemar J.Cox, Natasa Milic-Frayling, Ken Wood 185

Term Frequency Normalisation Tuning for BM25 and DFR Models
Ben He, Iadh Ounis 200

Information Retrieval Models (II)

Improving the Context-Based Influence Diagram Model for Structured Document Retrieval: Removing Topological Restrictions and Adding New Evaluation Methods
Luis M. de Campos, Juan M. Fernández-Luna, Juan F. Huete 215

Knowing-Aboutness: Question-Answering Using a Logic-Based Framework
Terence Clifton, William Teahan 230

Modified LSI Model for Efficient Search by Metric Access Methods
Tomáš Skopal, Pavel Moravec 245

PIRE: An Extensible IR Engine Based on Probabilistic Datalog
Henrik Nottelmann 260

Text Classification and Fusion

Data Fusion with Correlation Weights
Shengli Wu, Sally McClean 275

Using Restrictive Classification and Meta Classification for Junk Elimination
Stefan Siersdorfer, Gerhard Weikum 287

On Compression-Based Text Classification
Yuval Marton, Ning Wu, Lisa Hellerstein 300

User Studies and Evaluation

Ontology as a Search-Tool: A Study of Real Users' Query Formulation With and Without Conceptual Support
Sari Suomela, Jaana Kekäläinen 315

An Analysis of Query Similarity in Collaborative Web Search
Evelyn Balfe, Barry Smyth 330

A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation
Cyril Goutte, Eric Gaussier 345

Exploring Cost-Effective Approaches to Human Evaluation of Search Engine Relevance
Kamal Ali, Chi-Chao Chang, Yunfang Juan 360

Information Retrieval Methods (II)

Document Identifier Reassignment Through Dimensionality Reduction
Roi Blanco, Álvaro Barreiro 375

Scalability Influence on Retrieval Models: An Experimental Methodology
Amélie Imafouo, Michel Beigbeder 388

The Role of Multi-word Units in Interactive Information Retrieval
Olga Vechtomova 403

Dictionary-Based CLIR Loses Highly Relevant Documents
Raija Lehtokangas, Heikki Keskustalo, Kalervo Järvelin 421

Multimedia Retrieval

Football Video Segmentation Based on Video Production Strategy
Reede Ren, Joemon M. Jose 433

Fractional Distance Measures for Content-Based Image Retrieval
Peter Howarth, Stefan Rüger 447

Combining Visual Semantics and Texture Characterizations for Precision-Oriented Automatic Image Retrieval
Mohammed Belkhatir 457

Web Information Retrieval

Applying Associative Relationship on the Clickthrough Data to Improve Web Search
Xue-Mei Jiang, Wen-Guan Song, Hua-Jun Zeng 475

Factors Affecting Web Page Similarity
Anastasios Tombros, Zeeshan Ali 487

Boosting Web Retrieval Through Query Operations
Gilad Mishne, Maarten de Rijke 502

Posters

Terrier Information Retrieval Platform
Iadh Ounis, Gianni Amati, Vassilis Plachouras, Ben He, Craig Macdonald, Douglas Johnson 517

Físréal: A Low Cost Terabyte Search Engine
Paul Ferguson, Cathal Gurrin, Peter Wilkins, Alan F. Smeaton 520

Query Formulation for Answer Projection
Gilad Mishne, Maarten de Rijke 523

Network Analysis for Distributed Information Retrieval Architectures
Fidel Cacheda, Victor Carneiro, Vassilis Plachouras, Iadh Ounis 527

SnapToTell: A Singapore Image Test Bed for Ubiquitous Information Access from Camera
Jean-Pierre Chevallet, Joo-Hwee Lim, Ramnath Vasudha 530

Acquisition of Translation Knowledge of Syntactically Ambiguous Named Entity
Takeshi Kutsumi, Takehiko Yoshimi, Katsunori Kotani, Ichiko Sata, Hitoshi Isahara 533

IR and OLAP in XML Document Warehouses
Juan M. Pérez, Torben Bach Pedersen, Rafael Berlanga, María J. Aramburu 536

Manipulating the Relevance Models of Existing Search Engines
Oisín Boydell, Cathal Gurrin, Alan F. Smeaton, Barry Smyth 540

Enhancing Web Search Result Lists Using Interaction Histories
Maurice Coyle, Barry Smyth 543

An Evaluation of Gisting in Mobile Search
Karen Church, Mark T. Keane, Barry Smyth 546

Video Shot Classification Using Lexical Context
Stéphane Ayache, Georges Quénot, Mbarek Charhad 549

Age Dependent Document Priors in Link Structure Analysis
Claudia Hauff, Leif Azzopardi 552

Improving Image Representation with Relevance Judgements from the Searchers
Liudmila V. Boldareva 555

Temporal Shot Clustering Analysis for Video Concept Detection
Dayong Ding, Le Chen, Bo Zhang 558

IRMAN: Software Framework for IR in Mobile Social Cyberspaces
Zia Syed, Fiona Walsh 561

Assigning Geographical Scopes To Web Pages
Bruno Martins, Marcirio Chaves, Mário J. Silva 564

AP-Based Borda Voting Method for Feature Extraction in TRECVID-2004
Le Chen, Dayong Ding, Dong Wang, Fuzong Lin, Bo Zhang 568

Author Index 571

A Probabilistic Logic for Information Retrieval

C.J. 'Keith' van Rijsbergen

Department of Computing Science, University of Glasgow,
Scotland, UK
keith@dcs.gla.ac.uk
http://www.dcs.gla.ac.uk/~keith

Abstract. One of the most important models for IR derives from the representation of documents and queries as vectors in a vector space. I will show how logic emerges from the geometry of such a vector space. As a consequence of looking at such a space in terms of states and observables I will show how an appropriate probability measure can be constructed on this space which may be the basis for a suitable probabilistic logic for information retrieval.

1 Introduction

Why another paper on logic? There is now a substantial literature on the application of logic to IR [1], so what new can be said about this topic? For one thing there is no unique logic that is suitable for reasoning in IR, but a labyrinth of possible logics. The field narrows somewhat if one insists that a logic combines naturally with a measure of probability. It narrows even further if both the logic and probability can be seen to respect the geometrical structure of the space of objects in which one intends carry out plausible inference.

To model IR we need logic, probability and similarity. Usually each is treated separately within any model. Is it possible to combine naturally all three within one framework? Or, can one find a way of looking at things that takes all three paradigms into account? The answer is, yes, and this paper is a description of how one might go about doing this.

2 What Is Needed?

To open the discussion we will start by presenting a small number of building blocks in terms of which such a framework can be constructed. These are *states* and *observables*. Objects are modelled by states, and the measurement of properties such as 'relevance' and 'aboutness' are modelled by the values that observables can have with a probability. It is important to realize that properties do not belong intrinsically to a state, but rather that the value of a property emerges as a result of an interactive measurement of that property. This is an essential change from the traditional way of viewing 'relevance' or 'aboutness' as belonging to an object. Given a state one can ask a question about that state, such as a simple two-valued question. This may be

D.E. Losada and J.M. Fernández-Luna (Eds.): ECIR 2005, LNCS 3408, pp. 1 – 6, 2005.

a Yes/No question, such as, is this document relevant or not, and its answer will be either Yes, or No, with a corresponding probability for each.

The method of representation is as follows. Documents will correspond to state vectors, queries will correspond to operators, and relevance will correspond to an operator as well. All this takes place in a Hilbert space, which for all practical purposes can be thought as a finite-dimensional vector space with complex scalars. Let me empasise that we have identified relevance and queries with observables.

Observables will correspond to Hermitian operators [2] which are represented by self-adjoint matrices. It is a theorem in Hilbert space that any Hermitian operator can be represented as a linear combination of simple operators, drawn from a set of projectors one corresponding to each eigenvalue of the operator, and combined linearly with the eigenvalues as weights. Thus if a query is now represented by a matrix instead of a vector then it can be resolved into a set of Yes/No question, each question weighted appropriately. For a detailed discussion of this, see [3].

2.1 Properties of Observables

That observables are sensibly represented by Hermitian operators in Hilbert space is a long story derived from their introduction into Quantum Mechanics (see [3]). The most important properties that they have are that they have real eigenvalues, and that the corresponding eigenvectors form an orthonormal basis of the space. Hence, it can be shown that a measurement of an observable gives as an outcome a real number, and the probability of the outcome is a function of where the eigenvectors are in the space. This gives us the possibility of a perspective, or a point of view, from which to observe the objects in the space. In practice this means that any document is indexed (with probability) with respect to each eigenvector of the matrix representing the query. The default case is where an observable, such as a query, is represented as a vector with respect to the same basis that indexes the documents.

3 Enter John von Neumann

In the early part of last century John von Neumann realized that to calculate the probability associated with an observation in Quantum Mechanics it was essential to have a geometry on the Hilbert space from which the probability could be derived.

> 'Essentially if a state of a system is given by one vector, the transition probability in another state is the inner product of the two which is the square of the angle between them.' [4]

Such an inner product is like the cosine correlation from which the probability of a value of an observable can then be calculated. The same is true for IR if we represent our objects in the way described above. But von Neumann did more, he and Birkhoff [5] wrote a, now seminal, paper showing how the geometry of the space also produced a logic - a non-classical logic no less. Simply put, the set of subspaces in Hilbert space form an ortholattice which is isomorphic to a non-classical logic just like a Boolean lattice is isomorphic to a Boolean logic.

This analogy between the representations of a Boolean logic and a non-Boolean logic is closer than one would intuitively suspect. Boole[6] in his Laws of Thought stated as defining equation for his logic: $x(1-x) = 0$ for propositional variables x. The subspaces of Hilbert space can be seen as propositions, and the projectors P onto them as propositional variables. That is, corresponding to each subspace is a unique (orthogonal) projection. These projectors are of course idempotent operators satisfying, $P^2=P$, which can be rewritten as $P^2- P = 0$, or $P(P-I) = 0$ not unlike Boole's defining equation. These projectors are like simple questions, which have only, Yes, or No as an answer. As mentioned above any observable can be expanded as a linear combination of these Yes/No questions.

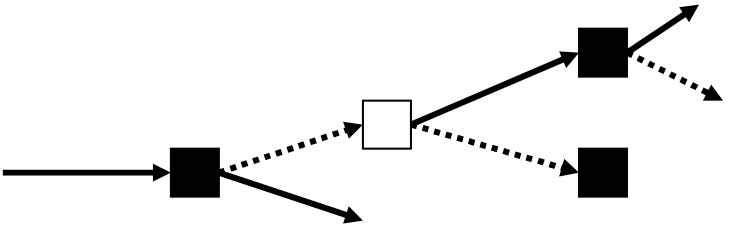
The set of projectors on a space – corresponding to the subspaces – in general, form a non-Boolean lattice. Once a point of view, the eigenbasis, has been chosen then the projectors on the eigenvectors can be combined to form a hierarchy of projectors. Any query is then expressible as a selection of these projectors.

4 What Is Different?

There is a famous example due to Wittgenstein [7] derived from Jastrow in which a drawing of the head of a duck-rabbit is shown. When presented with such a drawing, people will see it as either a rabbit or a duck. Of course once told of what it should be they will see it that way, and when shown it repeatedly in quick succession, the same decision is made. The above example, illustrates how 'rabbitness' or 'duckness' is not a property of the representation (figure). The seeing of one or the other emerges during the interaction.

The attribution of relevance or aboutness can be analysed in the same way. It is not that 'relevance' or 'aboutness' is a property of a document, it is rather that such a property emerges through the user interaction with a document mediated by his, or her, cognitive state.

In general the measurement of different observables will interact. Mathematically this comes down to assuming that the operators corresponding to the different observables do not commute. In IR this is quite natural, if one of the observables is relevance one would expect the outcome of a judgment of relevance followed by a judgment about contents, to be different from a sequence in reverse order simply because there is a cognitive state change between two such judgments.



A document for assessment comes in at the left to be judged as to whether it is about 'ducks' (black box), assuming it is not (dashed line) and is then judged for relevance (white box) and is considered relevant (solid line). During the process of

relevance judgment a user will have a cognitive state change, and when subsequently presented again with the same document for assessment as to whether it is about ducks or not, he may change his mind giving the possibility of either outcome. Thus the 'aboutness' assessment interacts with the 'relevance' assessment.

We are now quite used to this kind of interaction when considering counterfactual reasoning. For example, consider the following statements about a glass.

(1) will it break? -> yes
(2) is the floor made of rubber? -> yes
(3) will it break? -> no.

Examples are all very well but how does one formally capture this kind of interaction and reasoning? Fortunately, Hilbert space theory used to represent Quantum Mechanics can also be used to represent IR and will indeed give an appropriate method for handling non-commutative operators.

By choosing Hilbert space as a vehicle for representing objects in IR we are committing ourselves to a particular kind of logic. The logic associated with QM, is known as quantum logic and in general is non-distributative (see [8]). What breaks down in such a logic is the distribution law [9],

$$B \wedge (H \vee L) = (B \wedge H) \vee (B \wedge L)$$

5 Algebraic Considerations

One of the requirements for a reasonable logic is that it should have the usual connectives, especially an appropriate implication connective [10]. It can be shown quite readily that quantum logic does indeed have a sensible implication. The easiest way to do this is algebraically. Let us restrict our discussion to projection operators which correspond to propositions in quantum logic. Then if E and F are such operators, namely $E^2 = E$, and $F^2 = F$, then we can define in Hilbert space H,

$$[[E]] = \{x \mid Ex = x, x \varepsilon H\}$$
$$E \leq F \text{ if and only if } FE = E$$

The '$\leq$' is the natural ordering on the subspaces of H, or the equivalent projection operators. We can now define,

$$[[E \rightarrow F]] = \{x \mid FEx = Ex, x \varepsilon H\}$$

Then the semantics of $E \rightarrow F$ is given algebraically, and it can be shown that $E \rightarrow F$ is a projector[3]. With this definition of '$\rightarrow$' we end up with a full blown non-classical logic. The use of logic has been written about extensively, and we will not discuss this any further here, those interested should consult [1].

It is interesting that the implication connective defined algebraically above is the Stalnaker conditional. This implication was the basis for the probability kinematics developed for IR in [11]. There, a probability revision mechanism know as *imaging* was proposed. It is an open problem as to how imaging might be specified in Hilbert space. It may turn out to be a simple application of the following Theorem.

6 Enter Gleason

The purpose of this paper is to show how probability may be combined with logic to support plausible inference in IR. To complete the story we need one more piece of formal development, namely how to link probability in with the logic and geometry of Hilbert space. For this we need some more mathematics. But first an acknowledgment to Schrödinger whose interpretation of the state vector in QM foreshadowed precisely the introduction of probability.

> 'It [state vector] is now the means for predicting probability of measurement results. In it is embodied the momentarily-attained sum of theoretically based future expectation, somewhat as laid down in a catalogue'(for source, see [3]).

A possible reading of Schrödinger's remark is that the state vector encapsulates in it all the information for predicting the probabilities of measurement outcomes for any observable, like the possible future uses of a library catalogue. Another way of putting this is that the state vector induces a probability measure on the entire space by associating a probability with each subspace.

There is a famous theorem by Gleason[12] which gives an algorithm that specifies exactly how an special kind of operator will induce a probability measure on the space of objects, and conversely how any probability on the space can be capture algebraically by such an operator. The theorem goes as follows[13],

> Let μ be any measure on the closed subspaces of a separable (real or complex) Hilbert space H of dimension at least 3. There exists a positive self-adjoint operator T of trace class such that, for all closed subspaces L of H, $\mu(L) = tr(TP_L)$.

The technical definitions do not matter here, one can look them up in [13], what does matter is that in this theorem all three probability, logic and geometry are combined. The logic is given by the projectors P_L, the probability by μ, and the geometry by the trace function tr(). Tr(T) is defined as a sum of inner products, Σ $[e_i | Te_i]$, where e_i is any orthonormal basis for H. Notice that this is an existence theorem, it claims the existence of a unique T once the probability measure has been specified.

To appreciate the power of this result. Assume that an a priori probability measure has been specified by using the query as an operator T. Now imagine that this probability measure is revised in the light of some feedback information, then the revised probability measure implies the existence, according to Gleason, of an operator T' which represents the new probability measure. This is a form of query expansion, T being expanded into T'. Another illustration is the use of conditional information such as $E \rightarrow F$, remember that this corresponds to a subspace, and so there is a projector $P_{E \rightarrow F}$ corresponding to that subspace. This projector can enter into the probability calculation as specified in the theorem.

7 Conclusions

In this paper I have presented a very sketchy introduction to how Hilbert space theory combined with Gleason's Theorem can be used to combine logic, probability, and