A note from the new chairman
Jan J IJdens

How else could I start but by thanking Mark Dunlop for chairing the BCS-IRSG for so many years! Mark has been very successful at making the BCS-IRSG a great forum representing all IR researchers and practitioners in the UK. We are sad he is leaving, but it is great to see Mark is taking up an exciting new research post in Denmark. I am sure he will have a great time and he won't manage to stay away from the IRSG for too long! :-)

Apart from Mark's leaving, the 1999 IRSG colloquium also saw the election of one of the largest committees the IRSG has ever seen. I believe this is excellent news as it clearly shows many of our members are willing to take an active role in keeping the IRSG alive. This year's committee consists of people with a wide variety of backgrounds and experience; I am sure that together we can help the IRSG stay active and relevant to you.

Ian has asked me to briefly introduce myself. So here it goes:

After an MSc in computer science specialising in IR I spent nearly four years as a Research Assistant at the Robert Gordon University in Aberdeen, working on object-oriented IR application frameworks, and the use of probabilistic inference networks. In Aberdeen I was also involved in the lecturing on the MSc module in IR.

Since 1997 I am working in the Information Technology department at Sharp Laboratories of Europe in Oxford, where I hold the position of Senior Research Scientist. My current research interests include information management and retrieval in general, with specific interest in multimedia information management and cross-lingual IR.

As far as the BCS-IRSG is concerned, there are exciting times ahead! Over the next year or two a number of important issues will come to the foreground. First of all, we are reviewing BCS-IRSG membership. We want to start formulating what the IRSG is all about, what you are getting out of your membership, and how we can make sure the BCS-IRSG serves all members as well as possible. The second major issue for the next two years is the future of the BCS-IRSG Colloquium. Over the past few years it has developed from a pure UK colloquium into a truly European event, attracting speakers and delegates from many different countries. In 1997 the IRSG ventured outside the UK for the first time to hold a very successful colloquium in Grenoble, underlining its European spirit. Over the next years we want to formalise this by turning the BCS-IRSG Colloquium into a true European IR conference, hopefully under the umbrella of the BCS and CEPIS jointly. The challenge is to do this while at the same time making sure that the BCS-IRSG still serves all UK members (that's you!) in the best way possible. We will try to extend the activities of the BCS-IRSG in Britain, e.g. by organising more workshops and co-hosting/sponsoring meetings together with other organisations. I would like to conclude by saying that I personally would like to see as many members as possible being actively involved in the BCS-IRSG. I ask anybody who would like to have any input on what the IRSG could do for you, what we could/ should change and what the IRSG should be all about to send me an email (chair.irsg@bcs.org.uk).

I wish you all a great summer, and look forward to hearing from you soon!

Jan

Who's who ━━

IRSG Committee Contact List 1999 - 2000

Wondering who you should contact about what? Well, here's the current list of contacts.

B C S

Chair
Jan J IJdens
Sharp Laboratories of Europe
Oxford Science Park
Oxford OX4 4GA
Email: jan@sharp.co.uk (or chair.irsg @bcs.org.uk)

Vice-Chair
Mounia Lalmas
Department of Computing Science
Queen Mary and Westfield College
University of London
Email: mounia@dcs.qmw.ac.uk

Secretary
Jane Reid
Department of Computing Science
University of Glasgow
Email: jane@dcs.gla.ac.uk

Treasurer
Margaret Graham
Institute for Image Data Research
University of Northumbria
at Newcastle
Email: margaret.graham@unn.ac.uk

Colloquium 2000
Stephen Robertson
Microsoft Research Ltd
St George House, 1 Guildhall Street,
Cambridge CB2 3NH, U.K.
Email: ser@microsoft.com

One-day events
Mounia Lalmas
as above

Ordinary members
John Davies
Information Access Research.
BT Laboratories
Email: john.davies@bt-sys.bt.co.uk

Ayse S Goker-Arslan
School of Computer and Mathematical Sciences
The Robert Gordon University
Email: asga@scms.rgu.ac.uk

David Harper
School of Computer and Mathematical Sciences
The Robert Gordon University
Email: d.harper@scms.rgu.ac.uk

Monica Landoni
Department of Information Science
University of Strathclyde
Email: monica@dis.strath.ac.uk

John Lindsay
School of Information Systems
Kingston University
Email: lindsay@kingston.ac.uk

Andrew MacFarlane
Department. Of Information Science
City University
Email: andym@soi.city.ac.uk

Kerry Rodden
Computer Laboratory
University of Cambridge
Email: Kerry.Rodden@cl.cam.ac.uk

Tony Rose
Canon Research Centre Europe Ltd
Guildford
Email: tgr@cre.canon.co.uk

Steve Wade
School of Computing and Mathematics
University of Huddersfield
Email: s.j.wade@hud.ac.uk

Informer Team
Jon Ritchie
Department of Computing Science
University of Glasgow
Email: jon@dcs.gla.ac.uk

Ian Ruthven
Department of Computing Science
University of Glasgow
Email: igr@dcs.gla.ac.uk

Committee update

Careful readers of the committee contact list (or those who attended the last IRSG AGM) will note the ever-changing face(s) of our committee.

The committee was elected in a block-vote with the named positions (chair, vice-chair, etc) being selected by the new committee. Jan-Jap IJdens was installed as new chairperson and Mounia Lalmas has taken over the vice-chair position. Grateful thanks are due, from all of the IRSG, to Mark Dunlop and John Lindsay for their sterling work in these positions for longer than some of us can remember.

New members to the new committee include, Ayse Goker-Arslan, Andrew MacFarlane, Tony Rose and Stephen Wade. Stephen Robertson joined the committee as the colloquium chair for the 2000 Colloquium to be held in Cambridge, and Kerry Rodden was appointed as member with special interest in student affairs.

Other snippets

One-day events are featuring heavily on the horizon with a one-day meeting on Artificial intelligence and Information retrieval (organized jointly by the IEE and the IRSG) to be held in November, the Text Retrieval Conference (not TREC, the other one - organized jointly by the IRSG and the Institute of Information Scientists) planned for October and a one-day workshop on relevance feedback to be held in September *http://www.dcs.gla.ac.uk/rf/*. More info in the next Informer on the role of one-day events in the IRSG's activities (and how to run one!)

The proceedings from this year's colloquium will be published in the eWiC series, *http://www.ewic.org.uk/*

and, of course, the membership fees debate rages on....

Mira '99 Report

Report on Mira '99 conference
Glasgow - 14-16 April 1999

http://www.dcs.gla.ac.uk/mira/workshops/ conference

Mira was a EU-funded Working Group, initiated to investigate IR evaluation techniques in the current environment of interactive, and to a lesser extent, multimedia IR systems. The working group was comprised of fourteen core European groups with regular outside participants such as Rayal Fidel (U. Washington), Annelise Mark Petjerson (Risoe) and Stefano Mizzaro (U. Udine).

Mira set itself the task of developing evaluation methods that take into account the fact that

* IR techniques are beginning to be used in complex goal and task oriented systems whose main objectives are not just the retrieval of information.

* New original research in IR is being blocked or hampered by the lack of a broader framework for evaluation. This lack being highlighted by the difficulty of evaluating new media, the role of interaction, the change from static to dynamic evaluation and others.

The Mira '99 conference was the last workshop run by Mira and took the form of an open conference with keynote speakers, papers, group activities and panels.

In this short, personal, view of Mira, I will only outline the papers that form the proceedings of the conference (soon to be available through the electronic workshops in computing series, *http://www.ewic.org.uk/*. The proceedings consists of three types of papers: invited keynotes, papers by authors who were asked to present at the conference and papers by authors whose papers were not presented at the conference. There was no quality distinction made between the presented/not presented papers, the choice being one of balance.

In *'The application of work tasks in connection with the evaluation of interactive information retrieval systems: empirical results'*, Pia Borlund and Peter Ingwersen of Royal School of Library and Information Science, Copenhagen, outline a series of experiments on obtaining relevance assessments.

Given that the methods of constructing test collections, such as those used in the Cranfield model and lately in TREC, do not adequately represent the experience of users in making assessments against a situational, dynamic information need, how can we generate relevance assessments within a controlled evaluation environment?

In previous work Borlund and Ingwersen proposed using simulated work task situation: a work task situation being a sort of 'cover story', or a description of a situation that would promote an information need. The simulated aspect - the work task situations are created by the evaluators not the users - is important here. As the work tasks are only simulated they can be applied to a number of users, giving experimental control. The potential difference between simulated and real information needs is a major theme of this paper.

In this paper, Borlund and Ingwersen describe a series of experiments designed to evaluate the applicability of this approach using relatively novice searchers, in a laboratory environment. Their main finding is that, based on a number of empirical attributes such as number of queries issued per work tasks, search time per task, number of relevance assessments per task, simulated work task situations can be used in place of real information needs. That is, from the analysis of the selected attributes, there is no significant difference between the searches performed on real and simulated information needs.

The importance of relevance assessments and relevance assessors was highlighted again in the paper *'Negotiating a multidimensional framework for relevance space'* by Silvia Gabrielli, University of Padova and Stefano Mizzaro of the University of Udine.

In this paper, they report on the results of a group activity, or demonstration, carried out at a previous Mira meeting. This demonstration was designed to investigate Mizzaro's framework which categorises relevance on a 3-dimensional scale:

*Information needs: either *real* (a lack of information that has not yet formed, even implicitly, into a need for information), *perceived* (an implicit representation of the information-requiring situation and the possible means to satisfy the need), *expressed* (some form of request for information), or *formalised* (the actual request put to an IR system)

*Information resources: - the *types* of resources that a user can access when experiencing a need. These can be either a *set of documents*, a single *document* (or information object, and so conceivably be an object that it is not an actual document, e.g. a summary), *metadata*, or a *surrogate*

*Information use context - the context in which the information is used. This dimension is constructed from three components: *topic*, *task* and *user attributes*, where *topic* refers to domain, e.g. 'evaluation in IR', *topic+task* gives more information about how the information is to be used, e.g. 'evaluation in IR for writing a report for the Informer', and *topic+task+user attributes* includes descriptions about the person with the information need, e.g. 'evaluation in IR for writing a report for the Informer, I'm in a hurry and the next issue is due

Mira '99 Report ━━━━━

out on Friday'

The framework does not only classify relevance as an abstract entity but can also be used to classify IR systems as to what type of information seeking they can/are able to support (perceived information needs using sets of documents with topic and task information, formalised information needs using metadata and topic information only?) This paper discusses revisions to the original framework, including the deletion of the original *Time* dimension and discusses the importance of the framework the user access to IR systems. It also discusses, briefly, experiments carried out to investigate the issues raised by some of the dimensions.

In a second paper, '*Measuring the agreement among relevance judges*', Mizzaro considers the issue of consistency between relevance assessors. The importance of this in relation to new features of test collections and evaluation (such as multimedia, partial relevance assessments) is discussed as is a review of consistency, or lack of consistency, of relevance assessments in IR.

Mizzaro outlines different types of possible judgements (binary, scalar, or weighted, total or partial ordering) and describes that problems that may arise using the different types of relevance judgements. He then proposes measures for calculating the agreement between judges and groups of judges that can overcome some of the difficulties he describes.

The invited contribution, '*Information seeking as explorative learning*' by G. Brajnik of the University of Udine, presents some of the fundamental problems of *using* IR systems and some fo the techniques suggested to evaluate the usability of systems, in particular *task models* of evaluation. These models concentrate on evaluating the worth of a system in

supporting how people perform information in relation to actual tasks.

Brajnik distinguishes two levels of interaction:

presentation level: the part of the system with which the user interacts

navigation level: the part of the system that controls the interaction, and through which the user navigates the information space.

He stakes a claim to the second, navigation, level being the most important for the success of systems and compares the notion of a navigation level with other approaches and models. He then outlines the factors that must be taken into account in evaluating task models in relation to navigation, e.g. predictability, consistency and customisability.

'*The perceived similarity of photos - seeking a solid basis for the evaluation of content-based retrieval algorithms*' by Eero Sormunen, M.Markkula, K.Jarvelin, University of Tampere, highlighted the difficulty of extending traditional IR evaluation methodologies to new media, in this case, images. Unlike a number of image retrieval systems, the Tampere approach is not based on evaluating retrieval of images based on textual descriptions of the image content but the difficulty of evaluating images based on automatically-indexed images.

Their approach to evaluating this kind of system is to construct a test collection to evaluate systems based on how well they retrieve images that are already known or found. This reasons for the latter restriction are in the paper but the former restriction - moving back to test collections when others are moving away from them - is motivated by the argument that automatically indexed image systems are so primitive, there is a lack of real users and performance characteristics may not guarantee realistic system

behaviour necessary for user evaluation.

Their approach was test in a prototype collection construction exercise, using journalists who are experienced in selecting photographs to illustrate news stories. The use of a standard test collection is mollified by having the relevance assessors based their assessments on a form of simulated work task (as in Borlund and Ingwersen). In Tampere paper the tasks were varied in order to investigate what attributes the journalists were using in selecting photos.

The paper discusses in detail findings from this exercise, in particular the context of the photos and task were important. That context of task (e.g. layout of page affecting the choice of image) is not surprising, but that the context of the image became important - journalists rejected similar images because of the text caption not the content of the image!

'The word-association methodology' by Marianne Lykke Nielsen and Peter Ingwersen, Royal School of Library and Information Science, Copenhagen, investigates the use of the word association test. This familiar test - an examiner presents a subject with a word, the subject has to respond with the first word which comes to mind- has been proposed as a method of capturing the colloquial vocabulary and mental associations of the users of IR systems.

This paper starts by outlining previous attempts to use the word association methodology in IR and highlights the difficulty of evaluating them. In particular, the nature of the test and difference between uncontrolled (completely free association) and controlled (subjects may only give words of certain classes as responses) make cross-comparative evaluations difficult.

Lykke and Ingwersen give a

Mira '99 Report

comprehensive and interesting review of the approaches to using the methodology and their corresponding evaluations. They also point to the lack of concentration in these evaluations to the use of the methodology in examining how well the methodology supports users in searching. To investigate this aspect, they developed an evaluation framework involving real users from a specific domain and application of simulated work tasks.

Michael Heine, in *'Time to Dump P and R'*, starts by taking a philosophical look at some of the basic issues involved in the more standard Cranfield-like evaluation methods, in particular challenging some assumptions surrounding the loaded terms 'information' and 'relevance'. Heine bases his approach to evaluation of IR systems strongly from the viewpoint that information cannot be seen s as an observable entity capable of being embedded in documents or 'information' objects but must be seen as an internal process.

He uses this position to propose a new paradigm - the searcher adaptation and response (SAR) paradigm, which reassess the basic Cranfield notions of recall and precision. Recall, under the traditional Cranfield model is removed and replaced by the number of relevant documents that *inform* the user. This step is motivated by the argument that as relevance (although he regards this as a meaningless term) can only be assessed dynamically we cannot evaluate effectiveness based on assuming a preordained set of relevance assessments (as in the traditional definition of recall)

He proposes a distinction between standard precision (Cran-precision) in which all documents were assessed against an original, pre-search state as regards conformity to statement and precision in which documents are assessed as to whether they inform the user, irrespective of the similarity to the original search request (SAR-precision).

Heine then describes how the two paradigms can be integrated into one combined paradigm.

'Can rule-based indexing support concept-based multimedia retrieval in digital libraries? Some experimental results' by Ulrich Thiel, Andre Everts, Barbara Lutes and Adelheit Stein, of GMD Darmstadt, described a sophisticated architecture for digital libraries that not only support a variety of access methods but a variety of support tools (such as thesaurii) and media types. They then outline previous work on indexing images, in particular their own system for automatically generating index descriptions based on an abductive rule-based systems.

In *'Toward a theoretical framework for IR evaluating in an information-seeking context'* Amanda Spink, University of North Texas and Tom Wilson, University of Sheffield proposed a new evaluation measure - the *Information Problem Shift*. Specifically this is measured as the change in an information seeking problem stage after IR interaction. This work has a strong basis in studies of information seeking, again reflecting Mira's grounding in the move of evaluation from static to dynamic information seeking and from laboratory to real-life searching.

In particular this work is based on a previous model, integrating information seeking and definitions of relevance. This model is composed of three dimensions:

time - movements during and between search episodes, including strategies, tactics, feedback and other actions/states that separate stages of searching

interactive search models - represented by interactive IR models

set of situated actions - these include actions, decision, judgements that occur throughout a search episode

These dimensions are discussed in detail in the paper. This provides a framework for measuring change in state or action of the user. The focus is on the information-seeker rather than the information of documents themselves.

The keynote paper, *'Process and outcome: On the evaluation of IR systems in the age of interaction, GUIs and multimedia'*, from Stephen Robertson, Microsoft Cambridge and City University. In this paper Robertson examined the notions of *process* and *outcome* in traditional IR evaluation setting and discusses their limitations in new IR evaluative environments.

Note 1: These descriptions are based on the draft proceedings, check with the final, eWiC, version for accuracy of content and titles

Note 2: This is a personal view, not a definitive one

Ian Ruthven
University of Glasgow

Lumis '99 Report ▬▬

Workshop on logical and uncertainty models for information systems (LUMIS '99)

The purpose of this workshop was to promote discussion and interaction among members of the Information Systems community; in particular among those members with research interests in logical and uncertainty models for the treatment of semi-structured and unstructured information. We believe that a very large part of the information that will be available in future will be of this nature. But what has uncertainty and logic to do with accessing and managing semi-structured and unstructured information?

Uncertainty plays a very important role in the representation, access, and retrieval of information. The representation of information objects is often uncertain. For example, the extraction of index terms from a document or a query to represent the document or the query information content is a highly uncertain process. The data describing the redness of a red object present in a picture stored in a multimedia database is subject to a certain degree of uncertainty too. Probability theory is one way of dealing with uncertainty, but there are other approaches, such as, fuzzy logic, Dempster-Shafer's theory of evidence,

imaging, neural networks, and so on.

Logic plays a very important role in the representation, access, and retrieval of information. Logic has proved over centuries to be a very powerful modelling and reasoning tool, providing a degree of formality and correctness that can be very useful for manipulating information objects. For the task of retrieving information, logic has been used to build models that provide a rich and uniform representation of information with the aim to improve retrieval effectiveness. Classical logic is one possible approach. Other approaches include belief revision, situation theory, possible world semantics, abductive logic, conceptual graphs, conditional logic, description logic, and so on.

The Information Systems community is made of people coming from different fields: theoretical computer science, databases, information retrieval, hypermedia, digital libraries, artificial intelligence, to mention just a few areas. In this workshop, we have succeeded in gathering a heterogeneous community. This was reflected in the papers presented in the workshop:

A logical formulation of the Boolean model and of weighted Boolean models, G. PASI, Instituto per le Tecnologie Informatiche Multimediali, Consiglio Nazionale delle Ricerche

A geometrical view of relevance effectiveness in information Retrieval, S. DOMINICH, Department of Computer Science, University of Veszprem

Text passage classification using supervised learning, Y. BI, Faculty of Informatics, University of Ulster and F. MURTAGH, Department of Computer Science, The Queen's University

A logical approach to query reformulation motivated from belief change, G. AMATI, Department of

Computing Science, University of Glasgow and P.D. BRUZA, Distributed Systems Technology Centre, University of Queensland

Some uses of fuzzy logic in multimedia databases querying, D. DUBOIS, H. PRADE and F. SEDES, Institut de Recherche en Informatique (IRIT) - CNRS, Universite Paul Sabatier

Measuring the effects of and, and not, and or operators in document retrieval systems using directed line segments, M.H. HEINE, Department of Information and Library Management, University of Northumbria

Two applications of probabilistic argumentation systems to information retrieval, J. PICARD, Institut Interfacultaire d'Informatique, Universite de Neuchatel

To allow for maximum interaction, each presentation was scheduled for 40mn including questions. An hour discussion allowed further interaction between the workshop participants. Following the tradition of previous workshops on Information Retrieval, Uncertainty and Logics held in 1995 and 1996, the exchange of ideas was very stimulating.

Organising committee

Fabio Crestani - International Computer Science Institute, Berkeley, USA

Mounia Lalmas - Queen Mary & Westfield College, England

Mounia Lalmas
Queen Mary and Westfield College

SIGIR '99

University of California at Berkeley

August 15-19, 1999

http://www.sims.berkeley.edu/conferences/sigir99/

Preliminary program

Tutorials

The Tutorials are scheduled for Sunday August 15 and will occupy half a day in the morning or afternoon. The accepted tutorials appear below; tutorial descriptions and instructor biographies are now available.

Introduction to Information Retrieval.
B. Croft

Introduction to Machine Learning for IR.
D. Lewis and Y. Singer

User interface design for information retrieval in e-commerce.
A. Marcus

Content Based Image Retrieval.
Manmatha and Ravela

Multilingual Information Access.
P. Schauble and J. Klavans

Implementation of High-Performance IR Systems.
A. Moffat and J. Zobel

Medical Informatics. **W. Hersh**

Inside Internet Search Engines.
J. Pedersen and Chang

Automated Text Summarization.
E. Hovy, Lin and D. Marcu

Theory and practice in text retrieval system evaluation.
C. Buckley and E. Voorhees

Advanced machine learning techniques for IR.
Y. Singer and D. Lewis

Workshops

Most Workshops are scheduled for Thursday August 19 and will occupy one full day. The accepted workshops appear below; workshop descriptions, due dates, and organizers are now available. One workshop is joint with

DL 99 and will be held on Saturday August 14. The workshops are:

Multilingual Information Discovery and Access (MIDAS)

Customized Information Delivery

Recommender Systems: Algorithms and Evaluation

Evaluation of Web Document Retrieval

Multimedia Indexing and Retrieval

Exploratory Workshop on Music Information Retrieval

Accepted Full Papers

The order of sessions does not necessarily reflect that of the final program.

User Interaction

Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces. Marc Sebrechts: The Catholic University of America, Joanna Vasilakis: The Catholic University of America, Michael Miller: The Catholic University of America, John Cugini: NIST; Sharon Laskowski: NIST

User revealment - a comparison of initial queries and ensuing question development in online searching and in human reference interactions. Ragnar Nordlie: Oslo College

From Reading to Retrieval: Freeform Ink Annotations as Queries. Gene Golovchinsky: FX Palo Alto Laboratory, Inc., Morgan N. Price: FX Palo Alto Laboratory, Inc., Bill N. Schilit: FX Palo Alto Laboratory, Inc.

Speech IR & Text Categorization

SCAN: Designing and evaluating user interfaces to support retrieval from speech archives. Steve Whittaker: ATT Labs research, Julia Hirschberg: ATT Labs research; john choi: ATT Labs research, Don Hindle: ATT labs research; Fernando Pereira: ATT Labs research, Amit Singhal: ATT Labs research

Document Expansion for Speech Retrieval. Amit Singhal: ATT Labs-Research; Fernando Pereira: ATT Labs-Research

A Re-examination of text categorization methods. Yiming Yang: School of Computer Science, CMU, Xin Liu: School of Computer Science

Distributed Retrieval

Comparing the Performance of Database Selection Algorithms. James French: University of Virginia, Allison Powell: University of Virginia, Jamie Callan: CIIR, University of Massachusetts, Charles Viles: SILS, Univ. North Carolina, Chapel Hill, Travis Emmitt: University of Virginia, Kevin Prey: University of Virginia, Yun Mou: CIIR, University of Massachusetts

A Probabilistic Solution to the Selection and Fusion Problem in Distributed Information Retrieval. Christoph Baumgarten: Dresden Univ. of Technology

Cluster-based Language Models For Distributed Retrieval. Jinxi Xu: CIIR, University of Massachusetts, Amherst, W. Bruce Croft: CIIR, UMass Amherst

LSI & Theory

Probabilistic Latent Semantic Indexing. Thomas Hofmann: UC Berkeley

A similarity-based probability model for Latent Semantic Indexing .

Chris Ding: Lawrence Berkeley Nat'l Lab/ U. California

Using a Belief Revision Operator for Document Ranking in Extended Boolean Models. David E. Losada: University of A Coruña, Alvaro Barreiro: University of A Coruña

Language Analysis

Cross-Language Information Retrieval based on Parallel Texts and Automatic Mining of Parallel Texts from the Web. Jian-Yun Nie: Universite de Montreal

A New Statistical Formula for Chinese Text Segmentation Incorporating Contextual Information. Christopher Khoo: Nanyang Technological University, Yubin Dai: Nanyang Technological University, Singapore, Teck Ee Loh: Data Storage Institute, Singapore

Information Retrieval Based on Context Distance and Morphology. Hongyan Jing: Columbia University, Evelyne Tzoukermann: Bell Laboratories - Lucent Technology

Summarization & Sentence Selection

Summarizing Text Documents: Sentence Selection and Evaluation Metrics. Jade Goldstein: Language Technologies Institute, Mark Kantrowitz: Just Research; Vibhu Mittal: Just Research, Jaime Carbonell: Language Technologies Institute

The Decomposition of Human-Written Summary Sentences. Hongyan Jing: Columbia University, Kathleen McKeown: Columbia University

The automatic construction of large-scale corpora for summarization research. Daniel Marcu: Information Sciences Institute

Efficiency & Scalability

Partial Replica Selection Based on Relevance for Information Retrieval. Zhihong Lu: University of Massachusetts at Amherst. Kathryn McKinley: University of Massachusetts at Amherst

Efficient Distributed Algorithms to Build Inverted Files. Berthier Ribeiro-Neto: UFMG, Nivio Ziviani: UFMG; Edleno Moura: UFMG; Marden Neuber: UFMG

Efficient Document Presentation with a Locality-Based Similarity Heuristic. Owen de Kretser: University of Melbourne, Alistair Moffat: University of Melbourne

Interactive Use of Phrases

Phrase Recognition and Expansion for Short, Precision-biased Queries based on a Query Log. Erika de Lima: University of Stuttgart, Jan Pedersen: Infoseek Corporation

The Paraphrase Search Assistant: Terminological Feedback for Iterative Information Seeking. Peter Anick: COMPAQ, Suresh Tipirneni: COMPAQ

Phrasier: a System for Interactive Document Retrieval Using Keyphrases. Steve Jones: University of Waikato, Mark Staveley: University of Waikato

Term Relationships

Combining Multiple Evidences from Different Types of Thesaurus for Query Expansion. Rila Mandala: Tokyo Insitute of Technology, Takenobu Tokunaga: Tokyo Institute of Technology, Hozumi Tanaka: Tokyo Institute of Technology

Context-Sensitive Vocabulary Mapping with a Spreading Activation Network. Jonghoon Lee: GSLIS, University of Illinois, David Dubin: GSLIS, University of Illinois

Deriving structure from texts. Mark Sanderson: CIIR, University of Massachusetts, W. Bruce Croft: CIIR, University of Massachusetts

Theoretical Models

A Hidden Markov Model Information Retrieval System. David Miller: BBN Technologies, Tim Leek: BBN Technologies; Richard Schwartz: BBN Technologies

Information retrieval as statistical translation. Adam Berger: Carnegie Mellon University, John Lafferty: Carnegie Mellon University

An Algorithmic Framework for Performing Collaborative Filtering. Jonathan L. Herlocker: University of Minnesota, Joseph A. Konstan: University of Minnesota, Al Borchers: University of Minnesota, John Riedl: University of Minnesota

MultiMedia Retrieval

Content-Based Retrieval using Heuristic Search. Dimitris Papadias: Dept. of Computer Science, Marios Mantzourogiannis: HKUST; Panos Kalnis: HKUST, Nikos Mamoulis: HKUST; Ishfaq Ahmad: HKUST

Content-Based Retrieval for Music Collections. Yuen-Hsien Tseng: Fu Jen Catholic University, Taiwan

Relevance Feedback Retrieval of Time Series Data. Eamonn Keogh: University of California, Irvine, Michael Pazzani: University of California, Irvine

---

Web search tools

The site at *http://www.searchtools.com/searchtools.html* gives masses of links to search engines (particularly for web searching), reviews and surveys of engines. Systems covered include : Java, Mac, Perl, Unix and Windows.It also contains a useful guide on how to choose, implement and maintain your a web search service.

Web growth and search engine coverage

Abstract: The amount of information on the World Wide Web is outpacing the coverage of the search engines indexing that information, researchers report. They say that may be the root of a phenomenon well known to Internet entrepreneurs: Not all Web sites are created equal.

According to the study, even the best search engine only keeps track of 16 percent of the Internet's Web pages. Collectively,the top 11 Web search tools only index 42 percent of the Web. Full article at *http://www.msnbc.com/news/287392.asp*

Workshop

Searching for information: Artificial intelligence and information retrieval approaches

A Colloquium organised by the Professional Group IEE A4 (Artificial Intelligence) and co-sponsored by the BCS IRSG group.

11-12 November, 1999. Glasgow

Aims and scope

The amount of available information is currently growing at an incredible rate; a particular example of this is the Internet. To use this information, whether for business or leisure purpose, we need techniques and tools to allow for fast, effective and efficient access to large amounts of stored information.

The field of information retrieval (IR), and more recently, the field of artificial intelligence (AI) have been looking at this problem. The IR field has developed successful methods to deal effectively with huge amounts of information, whereas the AI field has developed methods to learn the user's information needs, extract information from text, and represent the semantics of information.

However, both fields have suffered from the fact that each community is often unaware of the work of the other.

This event brings together the two fields of AI and IR, to allow for the sharing and combination of techniques, with the aim to improve the search process.

The event is composed of presentations given by leading researchers in AI and IR that are looking at developing efficient and effective access to large amount of stored information. The event will present both theoretical and applicative results on the use of IR and AI techniques to seek information. This event is a first contribution to bridge the gap between the two research communities.

Who should attend

The event will be of direct relevance to academics and post-graduate students working in the field, industrial-commercial researchers, and end-users of search systems. Participants will benefit from hearing about current theoretical and practical developments across a wide range of activities.

Contributions

To offer better interaction with participants, the event will contain poster sessions. These enable participants to present late-breaking results, work in progress, or work that is best presented interactively or graphically. If you are interested in presenting your work as a poster at this event, a description of 1,000 words or less of the work that will be covered in your poster should be submitted. Poster submissions should be sent to arrive by 1 September, 1999, to: Alison Cawsey, Department of Computing and Electrical Engineering, Heriot-Watt University, Edinburgh EH14 4AS. Email: *alison@cee.hw.ac.uk*

Registration

For a registration form (when available) please contact: IEE Events Office, Savoy Place, London WC2R 0BL, Tel: +44 (0)20 7344 5732/5733, Fax: +44 (0)20 7497 3633, Email: events@iee.org.uk. Delegates must arrange their own accommodation.

Organisers

Mounia Lalmas, Queen Mary & Westfield College, University of London, England

Alison Cawsey, Heriot-Watt University, Scotland

Keith van Rijsbergen, University of Glasgow, Scotland

Venue

The IEE building in Glasgow is centrally located with excellent facilities. See *http://www.iee.org.uk/ SEC/building.htm.*

Correspondence:

Direct correspondence, inquiries related to this event should be addressed to:

Mounia Lalmas, Department of Computer Science, Queen Mary & Westfield College, University of London, London E1 4NS, UK. Email: *mounia.lalmas@dcs.qmw.ac.uk*

Alison Cawsey, Department of Computing and Electrical Engineering, Heriot-Watt University, Edinburgh EH14 4AS. Email: *alison@cee.hw.ac.uk*

Programme(provisional)

Thursday 11th November

10.45 Opening and introduction

11.00 Yorick Wilks. *The issues of representations in AI, IR and NLP*

11.45 Dieter Fensel. *Applying AI to the web*

12.30 Lunch and poster session

14.00 John P Eakins. *How smart are current image retrieval techniques?*

14.45 Carole Goble *A picture representing triumph or similar: classification based navigation and retrieval for picture archives*

4:00 Theo Huibers. *Intelligent information retrieval agents*

4.45 Marc Moens. *Personalised information objects*

Friday 12th November

9.15 Opening and introduction

9.30 Karen Sparck-Jones. *IR lessons for AI*

11.30 Gianni Amati. *Learning by examples as relevance feedback, and relevance feedback as learning by examples*

12.15 Lunch and poster session

13.30 Keith van Rijsbergen. *Quantum Logic: A new paradigm for IR*

14.15 Stephen Robertson. *Probabilistic retrieval: Thresholding for automatic filtering*

15.30 Yves Chiaramella. *tba*

16:15: Discussion

17.30 Closing

Competition

Springer-*I*nforme*R*
Prize for best student paper in IR '98-'99

The Informer in collaboration with Springer-Verlag are pleased to announce the launch of the third annual competition for the best student paper in Information Retrieval. This is an open competition for any student in a European academic institution who has published a paper in a refereed journal/conference/workshop in the period 1st November 1998 - 1st November 1999. Springer-Verlag have kindly donated a prize of £200 worth of Springer-Verlag books (full catalogue at <http://www.springer.de/cgi-bin/SFgate?bookdealer=Springer>).

The winner of the competition will be invited to present their paper at the 22nd Annual BCSIRSG Colloquium to be held in Cambridge, hosted by Microsoft , and will receive £200 towards their travel and accomodation.

To enter
1. Send a postscript copy of your paper to *informer@dcs.gla.ac.uk*, including publication details (conference name and dates/journal name, volume and number)

Rules
1. The paper must have been **published** or **accepted for publication** in the period 1st Nov 98 - 1st Nov 99 .
2. The paper must have appeared in a *refereed* journal, conference or workshop proceedings and should have a significant information retrieval content.
3. The entrant must have been a student at a European institution (university, college, etc) at the time the paper was *written*.
4. The entrant must be the main or only author of the paper.
5. Each entrant can only submit one paper for consideration.

Closing Date: November 12 1999