



INFORMER



British Computer Society
Information Retrieval Specialist Group

Volume 7

Autumn 1997

ISSN 0950-4974

Oh oh!!!



Mark discovers the administration associated with 950 IRSG members

Welcome to the latest issue of the *Informer*. As you can see from our cover, Chairperson Mark is filled to the brim with joy to hear our latest membership figures.

Inside this fun-packed (sic) issue are two conference reports (RIAO '97, by Adrain Müller of IBM, and Digital Libraries '97, by Pia Borlund, of the Royal School of Librarianship, Denmark). We also have announcements on Sigir '98, workshops on Text Retrieval, Image Retrieval, and, of course, Le Colloquium '98.

Ian & Jon

Don't bin me – stop getting me.

If you no longer wish to receive the *Informer* then please save the group a little money and return this form. We'll then remove you from the mailing list and save you the hassle of binning it, the cost to us and the waste of paper.

Name:.....

Membership number:.....

See the envelope for your membership number.

Send to Specialist Groups, BCS HQ,
1 Sanford Street, Swindon SN1 1HJ



| | | | |
|-----------------|-------------------------|-------------------------------|-----------------------|
| Contents | Who's who 2 | Digital Libraries '97 8 | Conferences 13 |
| | JoSIT 3 | Sigir '98 9 | Text retrieval '97 16 |
| | Image retrieval 4 | Colloquium '98 10 | |
| | RIAO '97 5 | Book reviews 12 | |

Who's who

IRSG Committee Contact List 1996 - 1997

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

Chair

Mark Dunlop
Department of Computing Science
University of Glasgow
Tel: 0141 330 6035 Fax: 0141 330 4913
Email: mark@dcs.gla.ac.uk (or chair.irsg@bcs.org.uk)

Vice-Chair

John Lindsay
Kingston University
Tel: 0181 547 2000 ext 2699 Fax: 0181 547 7887
Email: lindsay@kingston.ac.uk

Secretary

Jane Reid
Department of Computing Science
University of Glasgow
Tel: 0141 330 5006 Fax: 0141 330 4913
Email: jane@dcs.gla.ac.uk

Treasurer

Jonathan Furner
School of Information and Media
The Robert Gordon University
Tel: 01224 283835 Fax: 01224 492608
Email: j.furner@rgu.ac.uk

Colloquium 98

Yves Chiaramella
CLIPS-IMAG Universite Joseph
Fourier 38041 Grenoble cedex 9
Grenoble France
Email: chiara@imag.imag.fr

One-day events

Mounia Lalmas
Department of Computing Science
University of Glasgow
Tel: 0141 330 5323 Fax: 0141 330 4913
Email: mounia@dcs.gla.ac.uk

David Harper

School of Computer and Mathematical
Sciences
The Robert Gordon University
Tel: 01224 262701 Fax: 01224 262727
Email: d.harper@scms.rgu.ac.uk



Clare Harvey

School of Computing and Information
Systems University of Sunderland
Tel: 0191 515 3508 Fax: 0191 515 2781
Email: clare.harvey@sunderland.ac.uk

Ruairi O'Donnell

Department of Information Science
Strathclyde University
Tel: 0141 552 4400 ext 3700/3096
Fax: 0141 553 1393
Email: rodonnell@dis.strath.ac.uk

Margaret Graham

Department of Information & Library
Management University of
Northumbria at Newcastle
Newcastle-Upon-Tyne, NE1 8ST
Email: margaret.graham@unn.ac.uk

Jan-Jaap IJdens

Sharp Laboratories Oxford Science
Park Oxford
Email: jan@sharp.co.uk

Informer Team

John Davies Information Access
Research BT Laboratories
Tel: 01473 605536 Fax: 01473 642459
Email: john.davies@bt-sys.bt.co.uk

Frances Johnson

Department of Library and Information
Studies Manchester Metropolitan
University
Tel: 0161 247 6156 Fax: 0161 247 6351
Email: F.Johnson@mmu.ac.uk

Jon Ritchie (co-editor)

Department of Computing Science
University of Glasgow
Tel: 0141 330 6854 Fax: 0141 330 4913
Email: jon@dcs.gla.ac.uk

Ian Ruthven (co-editor)

Department of Computing Science
University of Glasgow
Tel: 0141 330 6292 Fax: 0141 330 4913
Email: igr@dcs.gla.ac.uk

Competition

The Informer in collaboration with Springer-Verlag are pleased to announce the launch of the 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 1996 - 1st November 1997. Springer-Verlag have kindly donated a prize of £100 cash (or equivalent) or £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 20th Annual BCSIRSG Colloquium to be held in Grenoble. Their travel to and accommodation at the Colloquium will be paid for by the BCSIRSG.

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* in the period 1st Nov 96 - 1st Nov 97. This means actually published, not just accepted or submitted.

2. The paper must have appeared in a *refereed* journal, conference or workshop proceedings and should have a significant information retrieval or information science content.

2. The entrant must have been a student at a European institution (university, college, etc) at the time the paper was *written*.

3. The entrant must be the main or only author of the paper.

4. Each entrant can only submit one paper for consideration.

Announce

The Journal of Systems and Information Technology

The Journal provides an avenue for scholarly work that takes a systemic or holistic perspective in relation to areas such as information systems development, information technology and information systems management.

The Journal of Systems and Information Technology fosters primarily, although not exclusively, interpretive or qualitative research methods including ethnographic, genealogical, action research and case studies of various kinds. Research that uses quantitative methods, for example statistical surveys, will be suitable if they take a broad perspective of the problems and issues. This means very often that the social and political aspects will be considered as well as the technical.

We believe that the time is right for such a journal as there is a growing body of interpretive researchers in IS taking a systems view. The IT in JoSIT refers to development and use of information technology.

Target topics that will be relevant to JoSIT include but are not

limited to: IS/IT planning that takes a systemic approach

Innovative Soft Systems approaches used in information systems development

The integration of software and/or hardware technologies that provide holistic solutions to problems. For example, the integration of various aspects of software engineering paradigms.

Human Computer Interaction (HCI) problems tackled in an systemic or integrated way.

Research papers that promote the development of interpretive or qualitative research methodologies in relation to Information Systems through case studies.

Systems approaches in the management of information systems.

Holistic approaches in the development of technology policy and technology transfer.

Integrative methods of systems design.

Systems perspectives in IS/IT evaluation.

The use of metaphors as an integrative theme for aspects of IT/IS.

The influence of politics and culture on systems development and the use of information technology.

Book reviews: JoSIT will include book reviews and information concerning conferences in the holistic information systems field.

Publication and manuscript guidelines: Researchers as well as Information Systems Professionals are invited to submit papers for the Journal. All papers will undergo a blind refereeing process by at least three referees. Papers can be sent in hard or soft copy. Soft copies should be in Microsoft Word for MAC or PC format. The Journal will be published twice a year in **March** and **September**.

Further details are available upon request. All submissions must be original works which have not appeared elsewhere and which are not being considered for publication with another journal. As the reviewing process will be conducted anonymously, please leave your

name(s) off the manuscript. People are encouraged to send their papers either in by email or hard copy form to:

Craig Standing Department of Information Systems, Edith Cowan University, Joondalup, Western Australia 6027. Email: c.standing@cowan.edu.au Fax: 61 08 9400 5633 Tel: 61 08 9400 5545

Chief Editor:

Craig Standing, *Edith Cowan University, Western Australia*

Members of the editorial board:

- Rakesh Agrawal, *University of Western Sydney, Australia*
- Robert Flood, *University of Hull, UK*
- James Alleman, *University of Colorado, USA*
- Duncan Langford, *University of Kent, UK*
- Teodosio Perez, *Amaral Universidad Complutense, Madrid, Spain*
- Chantal Morley, *Institute National des Telecommunications, Evry, France*
- Richard Baskerville, *Binghamton University, N.Y., USA*
- Michael Myers, *University of Auckland, NZ*
- Paul Beynon-Davies, *University of Glamorgan, Wales, UK*.
- Dewald Roode, *University of Pretoria, S.A.*
- Subhash Bhatnagar, *Indian Institute of Management*
- Ross Smith Swinburne, *University of Technology, Australia*
- Bill Doolin, *University of Waikato, NZ*
- Hazel Suchard, *Australian Catholic University*
- Keith Ellis, *University of Humberside, UK*

Workshop

The Challenge of Image Retrieval

A Symposium and Workshop on Image Retrieval

February 5-6, 1998, Newcastle upon Tyne, United Kingdom.

Organizing Committee:

Mark Dunlop, *University of Glasgow*
John Eakins, *University of Northumbria at Newcastle*
Peter Enser, *University of Brighton*
Margaret Graham, *University of Northumbria at Newcastle*
David Harper (Chair), *Robert Gordon University, Aberdeen*
Joemon Jose, *Robert Gordon University, Aberdeen*
Martin Nail, *British Library Research and Innovation Centre*

Content-Based Image Retrieval is one of the most exciting and fastest-growing research areas in the field of multimedia technology. However, opportunities within the UK for the exchange of ideas between different groups of researchers, and between researchers and potential users of image retrieval systems have so far been limited. This two-day event is designed to bridge both of these gaps, bringing together researchers in information retrieval, database, computer vision and image processing communities, together with the users and providers of image and video libraries. Further, this event will enable us to identify the major research areas that need to be addressed by researchers developing image retrieval systems, and thus help to set the future agenda for UK research in this field.

The two-day event will consist of a research workshop on Thursday, 5 February, at which researchers from UK and mainland Europe will present their latest research findings, followed by a symposium on Friday 6 February, at which a distinguished panel of speakers, led by Prof. Ramesh K Jain of the University of California, San

Diego, will review the state of the art in content-based image retrieval, with emphasis both on technical developments and actual user requirements. The two events will be linked by a Conference Dinner on the Thursday evening.

Programme Details

This will be held in the Forte Post House Hotel, Newcastle. Delegates may register for either or both days of the conference. All two-day bookings will automatically include the conference dinner; one-day participants are also encouraged to attend the dinner in order to enlarge their range of contacts in the field.

Full programme details, together with details of costs and booking forms, will be circulated early in October.

Important dates:

Deadline for Submission: 24 November 1997
Notification of Acceptance: 5 January 1998
Camera-ready Papers due: 5 February 1997

Original research papers are solicited for the research workshop describing work in progress or completed work on any topic related to image and video storage and retrieval. Possible topics include, but are not limited to: Feature extraction and representation - Image search and browsing techniques - Query models, paradigms and languages for image retrieval -

Similarity retrieval of images and video - Semantic retrieval of images and video - Evaluation and user issues - Neural network techniques for image

retrieval - Image retrieval applications and systems - User interface issues - Database architectures for image retrieval - Image crawling and the net

Submission details:

Authors are asked to submit a 1500 word extended summary, in English, which clearly indicates the originality of their work and its contribution to the field of image retrieval. Summaries should be submitted to the Conference Chair, either electronically or on paper, to arrive no later than Monday 24 November 1997. Authors whose contributions are accepted for presentation will be asked to submit a complete version of their paper (no more than 5000 words) no later than the date of the workshop. All attendees at the workshop will receive a copy of the published proceedings.

Authors submitting on paper are asked to send 3 copies to Prof David Harper, Conference Chair, The Challenge of Image Retrieval, School of Computer and Mathematical Sciences, Robert Gordon University, Aberdeen AB25 1HG, together with a covering letter containing contact information. Authors will be notified of acceptance or rejection no later than 5 January 1998.

Provisional Programme

9.00 Registration and coffee
9.15 Welcome address
9.30 Research presentations
11:00 Tea/Poster session
11.30 Research presentations
12.30 Buffet lunch
1.30 Research presentations
3.00 Tea/Poster session
3.30 Research presentations
5.00 Workshop close

RIAO '97 Report



RIAO97 Conference is organized by the "Centre de Hautes Etudes Internationales d'Informatique Documentaires" (C.I.D), 36 bis rue Ballu, 75009 Paris France, Tel: (33) 01 42 85 04 75. There is no ISBN for the proceedings.

Report by Adrian A. Müller

IBM Germany, German Software Development Laboratory,
adrian_mueller@vnet.ibm.com

The series of conferences under the title RIAO has now been running since 1985 every third year in Europe and America. The conference is organized by the "Centre de Hautes Etudes Internationales d'Informatique Documentaire" (CID), Paris, France, along with various international affiliates. RIAO has the special feature of incorporating both scientific papers and innovative product demonstrations.

The 1997 topic was the computer-assisted information searching on the internet. However, more than 50 talks, 8 prototype and 12 product presentations also covered related topics, such as visualization, (multi-) linguistic techniques and information extraction, architectures of IR-systems and logical approaches to IR. Accompanying the conference was a workshop, which presented the results of the research cooperation between France and Quebec, plus panel sessions on juridical problems on the use of (public) data and a session on the multilingual aspects of TREC-6 and AMARYLLIS, the French TREC.

The conference opened with an invited talk presenting the GILS project. Brodie (GILS: an Evolving International Approach to Enhancing Resource Discovery) presented an overview of the GILS (Government Information Locator Service) project. GILS essentially is a profile (extended set of attributes) of the ANSI Z39.50 standard. It has evolved from a U.S. federal government initiative into an international standard. GILS has been mapped to the US MARC format and

to the Dublin Core metadata, as well. It is described at <http://www.usgs.gov:80/gils/index.html>. A survey of current resource description formats can be found at <http://www.ukoln.ac.uk/metadata/DESIRE/overview/>.

Information discovery

Two different aspects of information discovery were addressed in session 1.

Ackermann et al. described the DICA agent (Do-I-Care-Agent: Effective Social Discovery and Filtering on the Web). DICA monitors web pages and classifies changes against a user model. A collaboration mechanism allows users to train their agents by pointing at other agents. DICA can be found at <http://www.ics.uci.edu/CORPS/dica.html>

The work of Dkaki et al. (Mining Information in order to extract hidden and strategic information), which was presented by J. Mothe, concentrated on the classical data warehouse approach. The TETRALOGIE system is a workbench, consisting of statistical data mining and visualization modules, which allow a single user to post-process the content of a DBMS system. Mothe presented the GUI of the system, which helps in identifying correlations of data by several graphical means.

Information visualization

Session 2 continued the discussion of information visualization. Cockburn

and Jones discussed the navigation problems in web-(sub)spaces, which partly are caused by design faults (stack-based undo mechanisms of web browsers, restricted semantics of hyper-links in HTML, etc.) of typical web applications. In the appendix of their paper, they present an interesting collection of network visualization tools and layout screen-dumps like WebMap, MosaicG etc.

In the next talk, Wexelblatt (Footprints: History-Rich Web Browsing) investigated one visual metaphor in depth: the footprints of many people, using the same walkways on a campus. Depicting the typical navigation of many users on a web-space in the form of an active, Java-based graphical visualization, can guide new users to browse complex information spaces.

The last talk of this session did not fit very well into the visualization topic. Nie (Using Terminological Bases in Information Retrieval) was the first in a (hidden) series of talks, which formulated logical means to cope with the tasks of IR. He described a logical model, which expands the traditional thesaurus relations with a notion of uncertainty and uses statistical measures and user feedback to adapt the thesauri to a certain application area. The formal evaluation of two thesauri and six different adaptation methods showed the impact of terminological knowledge, which should be used both by the system and presented to the user.

Automatic Abstracting Repositories

Session 3 covered two talks about document abstraction and two talks on the access of WWW information.

Jang and Myaeng (Development of a Document Summarization System for Effective Information Services) describe a system to select candidate sentences of a document, based on a training phase and a new method of sentence selection. Their system computes multiple probability values based on several features (lexical and statistical) first and combines them with the Dempster-Shafer theory of evidence, secondly. The system has been applied to Korean texts.

The opposite approach was presented by A. Lehman (Automatic summarization on the Web) A System for Summarizing using Indication Fragments: RAFI). Designed for scientific and technical texts, RAFI relies on a manually coded dictionary of 5.600 words to perform a discourse analysis of a given text. We are left with the question as to what extent this design will expand to a broader coverage - and in what extent this will decrease the quality of the results.

Similar problems of hand-crafted information resources were in the work of Chidlovsky et al. (Towards Sophisticated Wrapping of Web-based Information Repositories). In the context of their Knowledgebroker project, they developed parsing tools to wrap html output (e.g., from a search engine) back into standard formats (e.g. BibTeX) for further processing. Obviously, one of the major problem they had to face was the rapid change of output formats for search engines of independent groups and companies.

Katz (Annotating the World Wide Web using Natural Language) presented the START project, which is available at <http://www.ai.mit.edu/projects/infolab/globe.html> since 1993. Web pages and queries are described and formulated by natural

language phrases, which are parsed and indexed as expressions of the form <subject relation object>. The key of START is to get everyone involved in describing their private WWW information resources.

Linguistic approaches for Information Retrieval

Arampatzis et al. (IRENA: Information Retrieval Engine based on Natural Language Analysis) take use of the AGFL (Affix grammar over a fine lattice) engine (see: <http://www.cs.kun.nl/agfl/>) to generate a lexical analyzer for English texts. Applying NLP co-occurrence techniques yields in high precision, but in low recall of results. The use of AGFL is continued in the PROFILE retrieval project (<http://vonkje.cogsci.kun.nl/~profile/>).

Pohlmann and Kraaij (The Effect of Syntactic Phrase Indexing on Retrieval Performance for Dutch Texts) systematically tested the influence of compound analysis (and separate indexing) on a collection of 60.000 Dutch texts and 36 queries. They concluded that the effectiveness of stemming, tagging and names detection can even be increased if one adds all head-modifiers of terms to the index.

This result was somewhat refined by the work of Mitra, Buckley et al. (An Analysis of Statistical and Syntactic Phrases). Buckley explained that the added effectiveness of phrase analysis seems to be diminishing if a good basic ranking scheme is being used. The situation changes slightly for determining the relative ranks of low ranked documents.

Multilingual approaches

Wechsler et al. (Multi-Language Text Indexing for Internet Retrieval) presented the SPIDER system, which indexes an intranet consisting of English, German, French and Italian documents. Queries are categorized

and processed for the corresponding indexing language by analyzing the stop-words in the query. Proper names and other language independent terms are treated separately to compute correct term weights.

In the talk given by Oard (Adaptive Filtering of Multilingual Document Streams) a combination of training (based on a bilingual corpus and the alignment of sentences) and automatic translation (based on the trained mapping of indexing vectors) was presented. Three variants of this cross-language retrieval method were tested on Spanish and English documents.

Hiemstra and de Joong (A Domain Specific Lexicon Acquisition Tool for Cross-Language Information Retrieval) also addressed the problem of cheap access to multi-lingual document collections. Their approach is a bilingual lexicon with probability information. Given a sequence of sentences in two languages, they compute the probability of a term-term translation with respect to the frequency of term-term occurrences in the training pairs.

IR architectures

The two sessions (6 and 9) on IRS Architecture hosted a variety of different engines, techniques and paradigms. In session 6,

Baldwin et al. (EAGLE: An Extensible Architecture for General Linguistic Engineering) discussed the cascaded linguistic processing steps (sentence detection, tagging, discourse analysis) and knowledge sources of their system, which was applied to different projects, e.g., an automatic management-take-over analysis for English texts, which took use of their co-reference identification module.

Case-based reasoning as a means to "parse" retrieval sessions on WWW information was introduced by Corvaisier et al. (Information Retrieval on the WWW using a decision making system). To find similar cases for an

actual situation in a retrieval session, their RADIX system must be informed about details of the user interaction. Since these data can not be accessed from commercial web browser, they decided to build their own browser first.

Another system on it's way to full operation mode was presented by Gudivada and Tolety (A Multiagent Architecture for Information Retrieval on the WWW). Multiple agents (control, user, search-result wrapper, ontology agents etc.) together build the HARNESS architecture. Unfortunately, many of these agents still need to be implemented.

Information Extraction

Session 7 actually consisted of two parts: a panel on TREC and AMARYLLIS, and three talks.

Experiences and challenges of the multilingual parts of TREC-5 and 6 were presented by Donna Harmann and Alan Smeaton. We like to refer the reader to the TREC Home-Page (<http://www-nlpir.nist.gov/TREC/>) for details, contact addresses and more. C Fluhr (AMARYLLIS: The french language TREC) gave an overview of the French AMARYLLIS conference, which uses French corpora of Le Mondes articles and the PASCAL database. The panel was quite successful, in the sense that - after the introduction given on TREC - AMARYLLIS was easily recognized to be a subset of the current multilingual (including French) TREC-6 design. However, the obvious question to the AMARYLLIS group, why they did not contact the TREC organizers in time to combine their efforts, remained open.

Black et al. (Integrated text categorization and information extraction using pattern matching and linguistic processing) gave an overview of a two-phase project on text categorization and information extraction. The COBALD demonstrator was re-used in the

FACILE project to categorize English texts according to a domain-specific ontology. The FACILE system increased the emphasis on the preprocessing of texts to overcome certain deficiencies of pattern-based analysis modules.

Given a hand-generated concept subsumption hierarchy, Zaiane et al. (On-line Resource Discovery Using Natural Language) map documents and queries to common concepts to improve precision as opposed to keyword-based indexing methods. They exploit several AI techniques and formalisms during their ongoing implementation works.

Gaizauskas and Robertson (Coupling IR and Information Extraction: A New Text Technology for Gathering Information from the Web) feed the output of a web search engine into an information extraction module. They applied their IE system VIE (Vanilla IE) with Excite generated summaries to improve the performance.

Document/relevance ranking

The talks presented in Session 8 on Document/Relevance Ranking illustrated the multi-purpose applicability of similarity measures in different contexts.

Mani and Bloedorn (Summarizing Similarities and Differences Among Related Documents) combined a tf-idf measure of words and phrases with a sentence and segment alignment procedure to pinpoint both similarities and differences between documents.

To overcome the problems of the typical short queries issued for web retrieval, Clarke (Relevance Ranking for One to Three Term Queries) introduced a relevance ranking named 'cover dense ranking'. This ranking is based on term proximity, combined with measures for the co-occurrence of terms. Their collection of queries, which they used for evaluation, can be copied from <ftp://plg.uwaterloo.ca/pub/mt/shortq>. They conclude that

cover dense ranking should be used as a special ranking technique for short queries only.

Cutting and Pedersen (Space Optimizations for Total Ranking) presented two algorithms for total ranking under given disk space limitations.

The last talk of session 8 was given by Rose and Wyard (A Similarity-based Agent for Internet Searching). Their similarity measure consisted (amongst others) of a combination of word frequency and word-n-gram statistics. Unfortunately, it remained unclear how this metric is related to the problem of web-crawling to search similar documents, which they used to evaluate their agent.

IRS architectures contd

Also the second session (#9) on IRS Architecture contained diverse topics.

The first talk, given by Belew (ARACHNID: Adaptive Retrieval Agents Choosing Heuristic Neighborhoods for Information Discovery) was withdrawn from publication, because it will be published in the proceeding of 1997 Machine Learning, Nashville. Belew promised to post his paper at <http://www.cse.ucsd.edu/~rik/RIAO97.ps>, but he didn't. Since Belew himself was not quite sure, whether the RIAO provided an appropriate audience for his ideas, I try to keep the summary simple: Keyword-eating pacmans get trained to cooperate on a hyper-media encyclopaedia and suffer from hunger, if they don't find the cookies (i.e., relevant documents). Afterwards, this wolf pack should be fit enough to survive on the web.

The talk of Mills et al. (Cobra: A new approach to IR system design) can best be summarized with his own executive summary: "COBRA is like a Database, but it is better!" COBRA is an IR framework build into an OODB Model, which separates data, content, IR paradigm and user models. On top of COBRA lies a Java based

GUI which supports query reformulation and tracing of query executions. COBRA is employed in several projects.

The second talk on logical information retrieval was presented by Müller and Everts (Interactive Image Retrieval By Means Of Abductive Inference). An abductive inference engines transforms query statements to low-level executable actions. Their retrieval framework MIRACLE (Multimedia Retrieval of Concepts in a Logical Environment) was applied to the domain of image retrieval. A training phase, based on manual classifications of selected images, provides a set of low-level axioms to map from a visual feature index (entropy, colors, fractal dimension etc) to descriptive items (cartoon, landscape, tree, ...). Thus, user can search for images on a semantical rich level, and the MIRACLE/Image system does the necessary translation to low-level features (eg, in a DBMS etc.) for them.

Query Reformulation

Boughanem and Soule-Dupuy (Query modification based on relevance backpropagation) feed positive and negative relevance feedback into a neural network. This network is constructed from term/document frequency information. By backpropagating query relevance judgements into the network, the ranking of documents could be improved.

The talk of Bruza and Dennis (Query Reformulation on the Internet: Empirical Data and the Hyperindex Engine) was number three in the logical IR paradigm. Bruza describes an internet search engine, that helps the user formulate their query by presenting a dynamically constructed hyperindex of potential search terms. Two variants of the system can be tested at <http://www.dstc.edu.au/cgi-bin/RDU/hib/hib> and <http://www.dstc.edu.au/cgi-bin/RDU/hotOIL/hotOIL>

contd page 15

Digital Libraries Report

The second ACM International Conference on Digital Libraries (DL '97),

**Philadelphia, USA,
July 23-26 1997.**

**Pia Borlund
The Royal School of
Library and Information
Science
Denmark.
pbj@db.dk**



The conference history of the ACM Digital Libraries conferences is rather short, so far. This conference was the fourth in the series of DL conferences, a follow up of the DL '94 held in College Station, Texas and DL '95 held in Austin, Texas. Both of these conferences were informal and served the purpose of getting together the researchers in this relatively young field. In 1996, ACM (the Association for Computing Machinery) adopted the series through two of it's Special Interest Groups, SIGIR (Information Retrieval) and SIGLINK (Hypertext). DL '96 was held in Bethesda, Maryland. The DL '97 marked the first affiliation of Digital Libraries with SIGIR, with SIGIR '97 taking place immediately after DL '97.

The DL '97 conference was used as the platform by the Americans to update each other on each of their ongoing Digital Libraries projects. An update which often was of an indirect and implicit nature and as a result not always very clear to, and informative for, the 'rest of the world'.

The conference covered various aspects, such as: how to handle images in multimedia environments;

different approaches to knowledge representation; several examples of 'user-oriented' agents; databases; system issues; and much more... The papers presented were very system oriented and at a high technical level, and lacked the conceptual discussions about what a Digital Library is and what it means to digitalise a library.

In continuation of the presentations, attention needs to be drawn to the committed plenary talk given by Law professor Pamela Samuelson. The talk was about new information technology and copyright issues — issues which are of relevance and importance to all of us with reference to the work carried out in our community.

For those of us, who attended the conference in order to find out about the concept of a Digital Library, we would learn that the concept: Digital Libraries, is the overall title for all the bits and pieces we normally think of as the various types of research carried out within the field of IR, and that the Digital Libraries '97 conference was more a SIGIR *junior* conference dedicated to a specific area of interest: Digital Libraries.



SIGIR'98

Department of Computer Science
The University of Melbourne
Parkville, Victoria 3052, Australia

Fax: +61 3 93481184

October 13, 1997

SIGIR'98

Melbourne, Australia

24-28 August, 1998

The Twenty-First Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, jointly hosted by the Departments of Computer Science at the Royal Melbourne Institute of Technology and the University of Melbourne, will be held in Melbourne, Australia, from the 24th to the 28th of August 1998.

The Call for Papers and other information relating to the conference is now available at <http://www.cs.mu.oz.au/sigir98/>.

Please register your name on the conference mailing list (for both physical and electronic mail) via the conference web page if you have an interest in submitting a paper to or attending SIGIR'98. If you do not have access to the electronic registration form, send your name and postal address to sigir98@cs.mu.oz.au or to SIGIR'98 at fax number +61 3 93481184 and you will be manually added to the mailing list.

Australia in particular and the Asia/Pacific region in general offers an enormous range of recreational opportunities. Most airlines flying to the south Pacific allow stopovers at other points for only nominal extra amounts. We encourage you to combine your attendance at SIGIR'98 with the holiday you have been meaning to take for several years now, but never managed to find time for.

Alistair Moffat,
Justin Zobel,
Chairs, SIGIR'98
Ross Wilkinson,
Program Chair, SIGIR'98

*Twenty-First ACM-SIGIR International Conference on
Research and Development in Information Retrieval,
Melbourne, Australia, 24-28 August, 1998.
<http://www.cs.mu.oz.au/sigir98/>*

Colloquium '98

Discovering new worlds of IR



<http://irsg.eu.org/colloq/cfp.html>

Grenoble, France
25-27 March 1998



colloq@irsg.eu.org

Call for Papers

Colloquium Chair: Prof. Yves Chiamarella

Programme Chair: Dr. Mark Dunlop

Call for papers

Papers describing work in progress or completed work are invited on any topic related to IR. Possible topics include, but are not limited to:

hypermedia - multimedia retrieval
- distributed network systems - natural language processing - formal models of IR - interface issues in IR - IR in library systems - networked IR - evaluation and testing - implementation issues in IR.

Authors are invited to submit a camera-ready copies of their paper, in English, to be received no later than 7 January 1998. Papers should contain at most 7500 words and should be formatted in accordance with Springer-Verlag's Electronic Workshop guidelines. The submission should include two copies of the paper: one anonymous copy for refereeing and one full copy for publication in the draft proceedings.

The organisation of the colloquium is planning to publish all or a selection of papers in the Electronic Workshops in Computing series, provided that the overall quality of papers is high enough. Accepted papers will be included as submitted in a draft

proceedings which will be distributed to all IRSG98 attendees. Authors will then have until 30 April 1998 to revise their paper in light of referees comments and feedback from the colloquium itself. The final version of papers will then be included in an e-WIC publication.

Please indicate if the paper is to be considered for the Best Student Paper Award. This Award requires that the first and primary author be a full-time student at time of submission.

Please note: papers should not be submitted to the electronic Workshops in Computing editors directly! Initial submission is as camera ready - not electronic.

Cepis sponsorship - application form

If you are looking for money to attend the IRSG '98 Colloquium then there is great news. CEPIS have kindly given us funding to allow a large number of researcher to attend the IRSG '98 Colloquium. If you wish to apply for this funding please fill out the form below. Successful applicants will receive up to £200.

The closing date for funding applications is Noon, Monday 23rd February 1998. Successful applicants will be notified within 7-14 days.

Criteria for Applications:

1. You must be a resident of a European Country
2. You are not receiving other funding to attend the IRSG Colloquium

Priority will be given to:

1. Young researchers in the field of Information Retrieval
2. New researchers in the field of Information Retrieval
3. Researchers presenting material at the Colloquium

Please fill out ALL sections of this application and send it to. Ruairi O'Donnell, <ruairi@dis.strath.ac.uk> The information you provide in this form will be used to evaluate the best applicants, so please be as informative as possible.

We are looking for three main pieces of information:

1. Contact Details

We need your name, address, email and phone number

2. Reference Contact an established Researcher

Only enter this information if you are a new/young researcher in the area of IR

3. Case for Attending the Colloquium

Give a detailed description of your research over the last 2 years Describe why you wish to attend the Colloquium

Book Reviews

The *Informer* would like to thank the *Computer Journal* for its permission to reproduce the Book Reviews featured in this issue.

Gary Marchionini

Information Seeking in Electronic Environments. Cambridge University Press, 1995. ISBN 0-521-44372-5. £30.00. US \$49.95. 224pp. hardbound.

This book aims to analyse how electronic technologies have changed the skills and strategies we use for managing information and how IT has augmented the 'fundamental human activity of information seeking'. It starts with an in-depth review of information seeking behaviour, electronic environments and information seeking as a task. It then carries on to discuss different strategies, established by observational research, which users have for searching and browsing large collections of information. Finally, Marchionini looks at the support IT gives for new information seeking behaviours to develop and how current information seeking behaviours will develop as searchers become more accustomed to performing on-line searches (for themselves on primary material and not through trained intermediaries on secondary material).

While the book clearly achieves the aims given above (even though I feel it focusses too much on boolean query systems), it is very much a review in nature and it is hard to see what direct lessons there are, for computing scientists, to take from it. But at a time when end user information retrieval (IR) systems are becoming more widespread and IR research is focusing more and more on providing usable interactive systems to end users, such a review is very helpful. Supporting the discussion are 14 pages of references covering a wide breadth of subjects. Much work on Human-Computer Interaction (the book is published as part of the Cambridge Series in HCI) is often criticised as being woolly. Unfortunately, Marchionini does not help counter this criticism. This is

partly because, outside its review of traditional behaviour, the book is simply too wide and shallow, introducing many issues almost in passing, and partly because of its discursive style.

In summary, although the book is superficial in places, a significant amount of research in user interaction with IR systems is covered. The book acts as a useful introduction and overview of the issues in an increasingly important area: the changing nature of commonly available IR systems and the increasing end-user focus in IR.

Mark D. Dunlop *University of Glasgow*

Bonnie A. Nardi (editor)

Context and Consciousness: Activity Theory and Human-Computer Interaction. The MIT Press, 1996. ISBN 0-262-14058-6. £ 33.95. 400pp. Hardbound.

Bonnie Nardi's (ed.) book is one of the most exciting, and at the same time frustrating, books to be published on the subject of human-computer interaction theory in recent years. Exciting because it clearly recognises a problem in contemporary research namely that 'There is a fundamental need for a theory of practice in human-computer interaction studies.' Exciting too because it promises that *activity theory*, a relatively recent import from the former USSR, may provide this long awaited theory. The book is frustrating, on the other hand, because it is difficult to glean from its pages enough detail about activity theory to confidently assess whether or not it meets this need. Reading the book in search of activity theory is rather like piecing together a jigsaw puzzle, but one where some pieces are missing,

and others don't quite fit together. For example, one fundamental concept in activity theory is that of the 'object' of an activity. But object is defined in several different ways in several different places by different chapter authors. Minimally, a glossary of important definitions could be provided. The difficulty of constructing such a glossary may reflect a fundamental problem in translating ideas expressed in the original Russian into English. Instead, Nardi and her co-authors prefer a more eclectic approach, in which varied definitions, with varied illustrations, implicitly provide an impression of what is meant by the terms. This approach, whilst perhaps less sterile than a list of definitions, may leave some readers wondering whether the concepts can be defined with sufficient precision to support a theory of practice in human-computer interaction. The book is structured into three sections, a basic introduction, some case studies of applying activity theory in practice, and three chapters exploring possible theoretical development. The early chapters introduce the reader to a rich set of ideas in a stimulating form. Nardi's own early chapter, which relates activity theory to distributed cognition and situated action theory, is extremely interesting, although I am sure practitioners in these related disciplines will have serious difficulties with Nardi's, at times, simplistic distinctions. In and of themselves, the case study chapters provide a wealth of interesting and innovative ideas on how to do contextual research. Quite apart from activity theory, the book is worth buying for these chapters. What is slightly disappointing is that it is not always clear how the case studies align to activity theory, and precisely what part activity theory played in their design. Once again the reader gets an impression of what activity theory

methods might look like, without being able to clearly identify its defining characteristics. It may be too much to expect one book to articulate all that is activity theory and this book is certainly not a 'how to do it' guide. However, Nardi makes no such claim for the book. If the authors' aims are treated cautiously as an effort to give the reader a taste of what activity theory has to offer and, minimally, to provide a language and a new way of looking at interaction in context, then we believe that the book achieves these aims. We participated in a reading group studying the book and it has certainly affected our outlook on the nature of contextual factors in human-computer interaction. Perhaps one of the best ways to read the book is to dip into chapters on an interest-driven basis. Don't expect to find activity-theoretic answers to your research problems, but do expect to find an interesting new way of looking at interaction in context.

Peter Wright and Andrew Dearden
HCI Group, University of York

Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi
Reasoning About Knowledge. MIT Press. 1995. ISBN 0-262-06162-7. £38.50. xiii+477pp. Hardbound.

Reasoning about knowledge is an exciting field of contemporary logic, placed as it is at the crossroads of such diverse disciplines as artificial intelligence (AI), economics, cryptography, pragmatics of natural language, database theory and distributed systems theory. It should not be confounded with that subfield of AI which goes under the name of 'knowledge representation and reasoning'. Although there is a substantial overlap of concerns between the two, and although both deal with the logical analysis of agents' knowledge, only the former presupposes an *explicit* representation of the notion of 'knowledge', usually

accomplished by means of operators fashioned on those of modal logic. Reasoning about knowledge is therefore that discipline concerned with analysing the implications of arguments such as 'Bob does not know whether he has mud on his forehead, but he knows that Alice knows it, and that she knows he does not know'. It cannot be claimed that this is a comprehensive handbook of reasoning about knowledge, as it is mainly devoted to the 'semantic' approach (a rival account of these notions, initiated by Montague and not covered in the present book, is based on the 'syntactic' treatment of epistemic modalities). However, this is by no means a substantial limitation, as the possible worlds approach is the one that has been most fruitful and proved most flexible in analysing knowledge and a number of applicatively interesting, related notions. The 11 chapters of the book embrace the 'epistemic' concepts of knowledge and belief of multiple agents (either natural or artificial), their relationships with time and action, their model- and proof-theoretic formalisations, their decidability and computational complexity properties, and their numerous applicative contexts. This work is presented within the unifying framework of Hintikka-Kripke's possible worlds semantics, presented in Chapter 2 in its simplest form and progressively extended throughout the book as more and more complex notions are analysed. No-one could be better qualified than these authors to write a textbook on this topic. Key contributors to the literature on reasoning about knowledge and its applications within AI, distributed systems and cryptography, they are to be credited for much of the recent development of this discipline and its growth in importance outside the mere philosophical logic circles. The book is written in a clear, terse style; the presentation and the mathematical arguments are both rigorous and detailed, and the close interplay of formal mathematical work and informal philosophical argument makes the book very pleasant to read. The book is as self-contained as a logic

textbook could be hoped to be, requiring only basic knowledge of logic methodology. In recent years, I have used some of this material (some had already been published in journal articles) while teaching undergraduate courses on logic and AI, and have found it of great pedagogical value, not least for its exemplary way of proceeding through the steps of a logical argument and formalisation. Each chapter is completed with a set of exercises and a comprehensive annotated bibliography. It is easy to foresee that this book will become a classic.

Fabrizio Sebastiani *Istituto di Elaborazione dell'Informazione Consiglio Nazionale delle Ricerche 56126 Pisa, Italy*

International Journal on Digital Libraries.

Special Issue on User Interfaces for Digital Libraries.

Guest Editors: Joseph A. Busch, (jbusch@getty.edu) and Isabel F. Cruz (ifc@cs.wpi.edu)

As digital libraries reach an increasingly wide variety of users that need to locate, extract, display, and summarize information, there is a growing need to provide effective user interfaces that can support these activities.

This special issue addresses the challenges in designing and building user interfaces for digital libraries. These challenges relate to the multimedia nature of the information, the amount of the information to be accessed and processed for presentation, the distinct functionalities expected by users ranging from naive to sophisticated, and the diversity of the applications and of the capabilities of the output devices.

We invite the submission to the special issue of manuscripts describing new and original research. The deadline for submission is December 15, 1997.

Conferences

Conference or Workshop information should be emailed to: informer@dcs.gla.ac.uk

Call for Papers. Second International Workshop CIA-98

COOPERATIVE INFORMATION AGENTS - Learning, Mobility and Electronic Commerce for Information Discovery in the Internet

3rd (Thu) - 8th (Tue) of July 1998 at the Agents' World Event 1998 Cite de Sciences - La Vilette, Paris (France).

Co-sponsored by the Center for Information Systems Integration and Evolution, Fairfax VA, USA. Daimler-Benz AG, Stuttgart, Germany.

In cooperation with the Special Interest Groups on Database Systems, Distributed Artificial Intelligence and Information Retrieval of the German Society for Computer Science (GI).

Invited Speakers

Tuomas Sandholm (Washington University, USA)
Robert Tolksdorf (Technical University of Berlin, Germany)
Katia Sycara (Carnegie Mellon University, USA)
Sharma Chakravarthy (University of Florida, USA)
Sandip Sen (University of Tulsa, USA)
Michael Huhns / Munindar Singh (University of South Carolina/ North Carolina State University, USA)
Edmund Durfee (University of Michigan, USA)

Workshop description

The research and application area of cooperative information agents is of

rapidly increasing importance. Information agents are computational software systems that have access to multiple, heterogeneous and geographically distributed information sources. The autonomous agents have to face up to the increasing complexity of modern information environments ranging from relatively simple in-house information systems, through large-scale multidatabase systems, to the visionary Infosphere in the Internet. Cooperative information agents work together in order to achieve private or global goals. One of their main tasks is an active search for relevant information in non-local domains on behalf of their users or other agents. This includes retrieving, analyzing, manipulating, and integrating information available from different information sources. The development of cooperative information agents requires expertise from several different research areas, especially AI, DAI, Databases, and CSCW. It is particularly important to investigate to what extent AI methods can be applied for information discovery by groups or teams of cooperative information agents in the Internet. This concerns, e.g., the use of efficient techniques from machine learning, evolutionary computing, and symbolic or numerical approaches for uncertain reasoning. Moreover, commercial aspects of information gathering in the Internet are becoming more and more relevant, e.g., agents are paid and have to pay for services. Thus, methods for rational, utility-based cooperation among the agents are needed. In addition, mobile information agents seems to be attractive for a flexible, and efficient information discovery in constrained environments.

The interdisciplinary CIA workshop series covers the whole

thematic range of cooperative information agents. Each workshop will focus on a few selected themes being of particular relevance and actuality. The CIA-98 workshop will build on the success of CIA-97 ('DAI meets Databases') and mainly focus on the themes 'learning', 'mobility' and 'electronic commerce' in the context of cooperative information discovery. Topics of interest include, but are not limited to:

Architectures of information agents. - Knowledge discovery and data mining in large-scale information systems. - Transparent access to heterogeneous information sources in the Internet. - Construction and use of ontologies for multiagent information gathering. - Learning, interaction and organization of multiagent systems for information discovery in changing environments - Communication among autonomous information agents. - Industrial applications of agent technology. - Mobile information agents in the Internet. - Collaborative information agents in distributed WWW applications. - Issues of programming cooperative information agents for the Internet. - Multiagent Systems and Geographical Databases. - Game-theoretic and other microeconomic principles for rational information agents. - Advanced protocols for negotiation and electronic commerce. - Information agents in electronic markets. - Security aspects for information discovery in the Internet.

Important Dates

- Deadline for Paper Submission February 6, 1998
- Notification of Acceptance/ Rejection March 23, 1998
- Deadline for camera-ready version April 10, 1998

Proceedings

The workshop proceedings including all accepted papers will be available for all registered participants at the workshop. The proceedings will be published by Springer as a volume in the LNAI series (Lecture Notes in Artificial Intelligence, subseries of LNCS).

Paper Submission

Authors are invited to submit papers describing both theoretical and practical work in the area of cooperative information agents. Topics of interest include, but are not limited to the ones listed above in the workshop description. Papers which describe ongoing research or provide an excellent surveying work are in particular welcome. All submitted papers will be refereed for quality, correctness, originality and relevance. Papers accepted or under review by other conferences, workshops or journals are not acceptable.

Preparation:

Please see the webpage at <http://www.informatik.tu-chemnitz.de/~klusch/cia98.html>

General Chair

Matthias Klusch (Technical University of Chemnitz, Germany)

Co-Chairs

Larry Kerschberg (George Mason University, USA)

Gerhard Weiss (Technical University of Munich, Germany)

CALL FOR PAPERS PODDP 98

Fourth International Workshop on Principles of Digital Document Processing. March 29—30, 1998 Saint Malo, France

PODDP 98 is the fourth in a series of international workshops that provide forums to discuss the modeling of systems that process digital documents using theories and techniques from, for example, computer science, mathematics, and psychology. PODDP 98 will take place in conjunction with the EP 98 Conference at the Palais des Congrès, Saint Malo, France. (PODDP 98 is, however, a workshop and not a conference.) The workshop will be held on two days, Sunday, March 29 and Monday, March 30, 1998. The charter of PODDP is deliberately ambitious and its scope broad. Indeed, we have added the adjective “Digital” to the series title to reflect the workshop’s emphasis on multimedia documents. The current state of digital document systems can be characterized as a plethora of tools without a clear articulation of unifying principles and underlying concepts. The practical and commercial impact of these tools — editors and formatters for many media (text, graphics, video, audio, animation), composition systems, digital libraries, WWW, word processing systems, structured editors, document management systems — is too pervasive and obvious to require further elaboration and emphasis. With the rapid development in hardware technology (processors, memory, and high bandwidth networks), however, the notion of a document and of document processing itself is undergoing a profound change. The growing use of multimedia has expanded our notions about content, scale, and dynamicity of documents. To address these changes, we hope to bring to bear theories and techniques developed by researchers in other areas of science, mathematics, engineering and the humanities (such as databases, formal specification languages and methodologies, optimization, workflow analysis, and user interface design.) PODDP is organized to promote a happy marriage between theories and techniques of digital documents and the systems that process them. PODDP provides an

ideal opportunity for discussion and information exchange between researchers who are grappling with problems in ANY area relating to digital documents. We invite researchers to submit papers that attempt to find a good balance between theory and practice. Papers should not exceed ten pages in length. One author of each accepted paper will be expected to present the paper at the workshop. Presentations will last about 25 minutes with some time for questions. We intend to produce the proceedings as a volume in the Springer-Verlag Lecture Notes in Computer Science series, as was done with the proceedings of PODP 96. Papers from PODP 92 and PODP 94 were published in special issues of the journal “Mathematical and Computer Modelling” (Pergamon Press).

Topics

Major topics of interest include, but are not limited to:

Automatic Generation of Documents - Digital Libraries
Digital Publishing Issues - Document Compilers
Document Databases - Document Dissemination
Document Management - Document Mining
Document Presentation - Document Recognition
Document Query Languages - Document Standards
Document Transformation - DSSSL
Hypertext and Hypermedia - HyTime
Information Retrieval - Intelligent Documents
Multimedia Documents - Multimedia Editing Systems
SGML, HTML, and XML - Theories of Documents
Theories of Media and Multimedia - World Wide Web

Paper submission

You are invited to submit SIX copies of a detailed abstract or a complete paper by December 12, 1997 to either program co-chair:

Charles Nicholas, Department of

Computer Science and Electrical Engineering, UMBC, Baltimore, MD 21250 USA

Derick Wood, Department of Computer Science, HKUST, Clear Water Bay, Kowloon, HONG KONG.

Each submission should have a cover page, which indicates the name, affiliation, address, electronic mail address, and telephone number of the contact author.

Submission by e-mail (to nicholas@cs.umbc.edu or to dwood@cs.ust.hk) is acceptable. In this case the paper should be in ASCII, LaTeX, or Postscript (in the latter case, it is crucial that there are line breaks).

Important dates

- Paper Submission Deadline: Friday, December 12, 1997
- Notification of Acceptance: Monday, February 2, 1998
- Copy Due for Preliminary Proceedings: Friday, February 27, 1998
- PODDP 98 Workshop: Sunday—Monday, March 29—30, 1998
- Final Version Due for Proceedings: Friday, May 1, 1998
- Camera-Ready Proceedings to Publisher: Monday, June 1, 1998

On-line information

Further information will be available over the Web at <http://www.cs.umbc.edu/conferences/poddp>, at <http://www.cs.ust.hk/~dwood/.poddp98>, and via the PODDP mailinglist. To subscribe to PODDP, send the message "subscribe PODDP" without quotes, in the body (not the subject line) of an email message to majordomo@cs.umbc.edu.

RIAO '97 contd

Grefenstette (SQLET: Short Query Linguistic Expansion Techniques, Palliating One-Word Queried by Providing Intermediate Structures to Text) utilizes NLP techniques (proper names extraction, noun phrase extractors) and affinities of terms and their syntactic contexts to guide users during query expansion. However, this approach seems completely corpus-based, i.e., it does not address the interactive aspects of the process of query formulation and result inspection.

Session 11 Information Filtering

The last talk on logical IR was presented by Amati and Crestani (Probabilistic Learning for Information Filtering). Although Amati's introduction to their probabilistic learning algorithm was very concise, the system demonstration of their ProFile prototype in the lobby demonstrated clearly the fast learning curve of their algorithm. ProFile filters news articles according to user profiles, which can be gathered from only a few positive and negative examples.

Chandrasekar and Srinivas (Using Syntactic Information in Document Filtering: A Comparative Study of Part-of-Speech tagging and Supertagging) defined 'Supertagging' to be a superset of conventional part-of-speech tagging. A word can belong to different supertags. In the process of parsing, the syntax-tree fragments of words are combined to clarify the actual supertag, an ambiguous word belongs to. The experiments showed, that supertagging improves retrieval quality (as opposed to POS tagging), and that it is significantly slower.

Querying recursive structures like SGML is an open issue for conventional DBMS systems. Marcoux and Seigny (Querying

Hierarchical Text and Acyclic Hypertext with Generalized Context-Free Grammars) introduced a formal data model and query language for those structures. They consider an efficient implementation of their parsing algorithm feasible for a subset of their model in their future work.

Information Extraction contd

Iwayama and Tokunaga (A Probabilistic Model of Passage Categorization) compare their approach of passage-based categorization with other techniques (e.g. Salton/Buckley, Mittendorf/Schaeuble, Hearst/Plaunt): It is based on probability theory, domain-independent (no NLP techniques) and separates summarization from categorization by taking into account the relationships between each passage and the target domain.

Hahn and Schnattinger (A Qualitative Growth Model for Real-World Text Knowledge Bases) describe a system which aims at the automatic learning of word meanings from context. Their clue is the combination of a basic terminological knowledge base (KB), linguistic criteria and reification of computed concept-hypotheses with the well-established KB. However, the quality of their approach is hard to estimate from the examples given.

Nakagawa (Extraction of Index Words from Manuals) describes an automatic extraction method for index terms, based on statistical relations between a compound word and its component simple nouns. Their algorithm can be parameterized for the number of extracted words by Rijsbergen's E-measure. Since their working domain were Japanese manuals, they first need to apply a morphological analyzer to identify word segments.

TEXT RETRIEVAL '97: NEW RESEARCH, NEW PRODUCTS

IIS/ BCS-IRSG EVENT

LONDON, BARBICAN CONFERENCE CENTRE, 4 NOVEMBER 1997

This One-day Conference is designed to provide a state-of-the-art review of theoretical and practical developments in text retrieval, with emphasis on effective text retrieval in a range of environments from in-house databases to the World Wide Web.

Programme

9:30 Registration and coffee

10:00 *Chairman's introduction*

Professor Peter Willett
(Department of Information
Studies, Sheffield University)

Research Session

10:15 *Overview of recent IR
research*

Professor David Harper
(School of Computer and
Mathematical Sciences, The
Robert Gordon University)

11:00 Coffee/Tea

11:30 *The TREC experiments*

Professor Steve Robertson
(Department of Information
Science, City University)

12:15 Questions and
discussion

12:45 Lunch

Product Session

14:00 *Text retrieval in-house
and on the Web - new PL
developments*

(Systematic Group or Personal
Library Software speaker)

14:30 *BT Intranet and filtering
technologies*

John Davies (Information
Access Research, British
Telecommunications plc)

15:00 Questions

15:15 Tea/Coffee

15:45 *Electronic text retrieval:
myths and practicalities*

Martin Earley (Global Recall
Ltd.)

16:15 *Overview of academic
and commercial IR
developments*

John Lindsay (School of
Information Systems, Kingston
University)

16:45 Questions and
discussion

17:00 Chairman's summing up

Cost:

**IIS and BCS IRSG
Members:** £144 + Vat

Non-members: £166 + Vat

Half day: £75 + Vat/ £85 +Vat

Booking address:

Institute of Information
Scientists

44-45 Museum Street

London WC1A 1LY

Fax: 0171 430 1270

email: iis@dial.pipex.com

The INFORMER

The *Informer* is published quarterly by the British Computer Society Information Retrieval Specialist Group.

Editorial Team: Ian Ruthven/Jon Ritchie, Department of Computing Science, University of Glasgow, Glasgow G12 8QQ. Tel: 0141 330 6292/6854 fax: 0141 330 4913. email: informer@dcs.gla.ac.uk

Advertising: Mark Dunlop, BCS IRSG Chair, Department of Computing Science, University of Glasgow, Glasgow G12 8QQ. email mark@dcs.gla.ac.uk