



MESSAGE FROM THE CHAIR

Dear BCS-IRSG Members,

Welcome to this issue of the *Informer* and welcome to new members of our specialist interest group. As Chair, I would like to take this opportunity to inform you about the various activities and events with which we have been involved.

Firstly, however, I would like to thank Jan Ijdens who as previous Chair ran the committee very efficiently and achieved a lot in the short period he was Chair. He also ensured a smooth transition by passing on relevant information so promptly. Jan has started a company in the Netherlands and we wish him success in his new venture. Also, I would like to thank Ian Ruthven for his efforts and contribution as the previous *Informer* Editor. We wish him the best in completing his PhD.

I am pleased to announce that Mark Girolami is our new *Informer* Editor, and Ali Shiri is the associate editor. The *Informer* used to be somewhat of an onerous task to produce particularly for the editors involved. So, I have decided to address this issue by broadening the input base for the content with specific responsibilities attached to committee members. However, we would also welcome any contributions you may have.

One recent development is the increasing Europeanisation of the activities of the BCS-IRSG. Currently, this is reflected in the choice of locations for our annual Colloquium (BCS-IRSG European Colloquium on IR Research) where we have established a pattern of holding the event outside the UK on alternate years. This year's Colloquium will be held in Darmstadt, Germany (April 4-6). We encourage you to submit



papers and support this event in our calendar. Further details of this and other events can be found inside this issue.

We are also looking at ways of ensuring greater European participation and involvement in our activities. To this extent, we would like to identify European representatives for different countries to help us keep in touch with the various European IR events and to provide input to committee activities from a broader perspective. Please get in touch with us if you are interested in such a role.

Another development at the international level is that the ACM-SIGIR executive has just recently selected Sheffield as the location for SIGIR 2004. The BCS-IRSG will be a co-sponsor of this event. Our congratulations to

Mark Sanderson (who will be general chair), Micheline Beaulieu and Peter Willett at Sheffield University for their successful bid for this prestigious international IR conference.

We would like to emphasise that the Committee greatly values suggestions and feedback from its members. Specialist interest groups like ourselves can only run as well as their members' input and enthusiasm enables. For example, we are keen for members to get involved with organising one-day events that address hot topics in IR and related areas. Please contact us if you have an idea or would like information on how to run such events.

Next, we are also working on extending our Web presence <http://irsg.eu.org/>. Our pages provide information on various events, courses, grants, and jobs. Whilst adding to the content we are considering the introduction of a discussion forum to enable more efficient interaction between our members.

Finally, I would like to wish you a happy and healthy 2001.

Ayşe Göker
Chair BCS-IRSG

CONTENTS

Message From The Chair	1
QuizXML	2
IRSG2000 Report	3
IR in the Netherlands	4
IR at Helsinki University	4
Call for Papers IRSG2001	5
Call for Papers SIGIR2001	6
Call for Papers WI-2001	7
Call for Papers JCDL2001	7
Job Opportunities	8
Whos Who	9

QUIZXML: AN XML SEARCH ENGINE

John Davies

Head of Advanced Business Applications BT

john.nj.davies@bt.com

This short article overviews research in progress into an XML-aware search engine for the WWW. As a meta-language, XML allows user communities from industry groups down to single users to develop an individual mark-up language that best suits their needs. XML will perhaps be most beneficial in applications that use semi-structured data, such as B2B applications and e-commerce. Microsoft, Ariba and CommerceOne and others have adopted XML for defining document interchange formats that lie at the heart of their B2B frameworks.

One of the potential benefits of XML is its ability to improve the accuracy of searches through the one billion documents now stored on intranets and the Internet. It is anticipated that intelligent exploitation of meta-information can deliver better retrieval performance as compared to today's HTML-based search engines.

An issue facing XML is of course the fact that so much of non-XML online information already exists. It is, as a recent Ovum report stated, a case of chicken and egg: "Until search engines include an XML capability, there is no reason to convert documents and vice versa." Nevertheless, considering the momentum behind its adoption and the number of XML-enabled products planned for release in the next 12 months, it seems highly likely that XML will become a widely used format.

Prior Art

Currently existing XML search engines allow the user to formulate queries using the mark-up vocabulary of the target documents. The user is therefore required to decide in advance which document type most likely contains the desired information. This is in contrast with the usual information seeking practice on the Internet where people search

for interesting information in all Internet documents. There will be many different XML document types on the Internet and the information seeker will not be aware of many of these. Hence it is impractical to expect the user to specify the XML document type prior to searching.

For example, the Internet XML search engine GoXML distinguishes itself from the above described type of XML search engines in that the user is not expected to be familiar with the mark-up vocabulary of potential target documents. Instead the user can start by entering a plain search query. GoXML then retrieves documents containing the search keywords and presents them to the user in a retrieval list.

GoXML also compiles a list of the tags that mark up document parts within which the search words were found. The user can then explore this list of hit tags and click on a particular tag in order to filter out those documents where the search keyword occurs in a part marked up by the selected tag.

QuizXML

QuizXML extends this functionality in that the user is also allowed to initiate his search by entering a plain search query, and then to filter out documents by selecting items from a list of hit tags, with the important addition that the tags are shown in context. Any tag is shown together with its superordinate and subordinate tags. Furthermore, when the user selects a tag from QuizXML's tag list, QuizXML adjusts the list of hit tags showing exactly those hit tags that occur in the filtered-out documents. The user is then free to further narrow down the search results by successively selecting tags from the simultaneously adjusted tag lists.

QuizXML indexes and searches over both XML and HTML documents, which is essential given the fact that the majority of information on the Internet is still currently encoded in HTML. The current version of QuizXML can be seen at www.quizxml.com. Further enhancements to QuizXML are in research and development at BT's Adastral Park Technology Centre. These include thesaurus-enhanced intelligent interpretation of tags; automatic ranking of tags,

based on previous search behaviour or on user profiles; and the semantic characterisation of tags based on the information contained within the tag across a set of documents. These are described in a little more detail below.

Thesaurus-enhanced Interpretation of Tags

XML tags are of course simply words, which can have exhibit properties such as synonymy and ambiguity. One author might tag part of a document as 'summary' while another might tag the same part of another document as 'precis'; or a section of one document might be about software agents and tagged 'agents' while in another document the same tag is used to tag a section about estate agents. Thesaurus enhancement can identify synonyms and help to disambiguate tags. The thesaurus used could be a general purpose one such as WordNet or a domain-specific one generated by clustering words occurring in a specific set of documents. Tags could be presented to the QuizXML user along with their synonyms to help clarify the context of the tag or, alternatively, related tags could be clustered together when presented in the search results.

Automatic Ranking of Tags

A user or group of users may be interested in particular types of information. QuizXML could infer this information either from a user profile (BT has technology in this area, previously patented); or from observing the tags which historically a user or group of users have shown interest in via their use of the tool. QuizXML's search results could then be ranked by the relevance of the tags to the user(s).

Semantic Characterisation of Tags

It is possible to look at the information associated with a given tag across a set of XML documents. This information can then be analysed, for example, by using ProSum to extract the key themes from the sections of information. The characterisation of the typical content of a given tag can then be presented back to the user, thus making clearer to the user what a given tag actually 'means'. QuizXML is part of an ongoing research project. As such, we would welcome both contributions of XML document collections in the public domain and comments on the current and planned system.

**THE BCS/IRSG 22ND ANNUAL
COLLOQUIUM ON
INFORMATION RETRIEVAL
RESEARCH**

*Sidney Sussex College, Cambridge,
England, 5-7 April 2000*

Ali Asghar Shiri
University of Strathclyde

This year about 100 people participated in the colloquium organised by the British Computer Society Information Retrieval Special Group (BCS/IRSG). Participants included university lecturers, students, researchers and information technology (IT) companies involved with information retrieval (IR) research and application. The gathering was sponsored by Microsoft Research Cambridge, AT&T and Dialog Corporation.

As this colloquium intends to reflect the information retrieval research and application undertaken throughout Europe, several delegates came from outside the UK, including participants from: Dublin City University and University College Dublin (Ireland), Universities of Nijmegen and Twente (the Netherlands), Laboratoire d'Informatique de Paris (France), Universities of Dortmund and Karlsruhe (Germany), Universidad Politecnica de Madrid (Spain), and the University of Tampere (Finland).

The majority of researchers, academics and students attending were from UK institutions of higher learning. Institutions represented included: City University London, Imperial College, Robert Gordon University, and the Universities of Cambridge, Edinburgh, Exeter, Glasgow, Huddersfield, Paisley, Sheffield, and Strathclyde.

During the two and a half day gathering 20 papers were presented on a wide range of information retrieval topics, ranging from natural language processing, user interaction, and topic detection, to indexing, probabilistic models, and visual navigation. The full list of titles was as follows:

- What we have learned and not learned from TREC
- Employing semantical issues in syntactical navigation

- An evaluation of linguistically-motivated indexing schemes
- Interactive searching behaviour: Okapi experiment for TREC-8
- Detecting session boundaries from web user logs
- Understanding information retrieval: a semiotic framework
- The design of topic tracking system
- Topic detection, a new application for lexical chaining
- Using term co-occurrence data for document indexing and retrieval
- Exploring the incorporation of acoustic information into term weights for spoken document
- Disambiguation strategies for cross-language information retrieval
- Dynamic selection of information retrieval algorithms
- Personalised retrieval for online recruitment services
- Integration of collocation statistics into the probabilistic retrieval model
- A probabilistic approach to Chinese information retrieval: theory and experiments
- Documents representations based on generative multivariate Bernoulli latent topic model
- Learning in information retrieval: a probabilistic differential approach
- A new effective approach for categorising web documents
- A visual information retrieval navigator
- Navigating n-dimensional information space with data and documents through view-based searching

Detailed information on these papers and their authors' affiliations can be found at: <http://www soi.city.ac.uk/~andym/colloq2000>. This summary report attempts to provide a succinct account of some of the most interesting papers which were presented.

Donna Harman as keynote speaker provided some interesting points on the lessons learned and not learned from the Text Retrieval Conference (TREC) which was started in 1992. She also elaborated on the evolutionary stages of the TREC related investigations and developments achieved throughout the last 8 years. The special tracks of the TREC conference like Spanish and Chinese track, Cross language track, routing

and filtering track, high precision track, interactive track, natural language processing track, query track, web track, database merging track were among the topics highlighted by Harman. She also raised some questions on evaluation of interaction in browsing by users, natural language processing, web test collection, cross language retrieval and the difficulties of translation which she believed to be of importance to the information retrieval research community.

Fowkes and Beaulieu in a paper on Interactive searching behaviour: Okapi experiment for TREC-8 examined searching behaviour in a highly interactive environment in the context of TREC-8 interactive track. Their study demonstrated that the effectiveness of relevance feedback is dependent on the characteristics of the topics. How the presentation of different interface features i.e. query expansion and relevance feedback, passage retrieval, displays of search results can be integrated into search process was among the research areas they suggested worth investigation. Hiemstra and Jong presented interesting development in the field of cross-language retrieval, summarising new findings related to strategies for linguistic disambiguation. Their research suggests that quality of search methods is more important than the quality of disambiguation methods and good retrieval methods are able to disambiguate translated queries implicitly during searching. The findings can be used as the basis for further investigation into several aspects of cross language retrieval.

At present, Web content classification and categorisation is considered to be an interesting area of Internet IR investigation. Klas and Fuhr reported on a new approach for categorising web documents which they term the megadocument approach. Using this approach a new document is classified according to a category which will be selected by a megadocument. Their findings also show this can be well suited for heterogeneous document collections.

Among the work on visualisation and information retrieval, Pollitt's work deserves mentioning. His paper, with some clarifying examples, showed how knowledge structures e.g. classification

systems could effectively be utilised in interface to help users rapidly identify required documents. Using an object-oriented approach to bridge the gap between document and data management systems, he has provided a view-based searching system through which users can navigate through and seek for information .

Understanding information retrieval : a semiotic framework by Murat Karamufoglu was among the papers which approached information retrieval understanding from a different perspective in this case through socially oriented semiotics. It was suggested that two distinct uses of IR systems can be postulated: for the transmission of knowledge from databases to users for didactics purposes but equally well for the production of new knowledge by inventing relationships between documents that constitute text retrieval databases.

In addition to the formal paper presentations, many informal discussions took place among delegates. The exchange of ideas and comments, which is an indispensable adjunct to such gatherings, was evident at this colloquium. As a Ph.D. student in the early stages of my research in the field of IR, I was grateful for the opportunity to talk to some of the leading thinkers in this field. It provided an excellent opportunity to receive comments and criticism on my proposed investigation and to highlight alternative research avenues for further investigation and contemplation.

By and large, the colloquium proved an excellent event for keeping abreast of recent IR research as well as providing hints as to the most exciting future research trends and tracks within the field.

The Informer gratefully acknowledges the Journal of Information Science for allowing us to reprint this report.

IR IN THE NETHERLANDS INTERIM REPORT

Kees Koster

All academic research in Computer Science in the Netherlands is organised into

four so-called Research Schools. Information Retrieval fits logically into SIKS the School for Information and Knowledge Systems. Part of the research in IR takes place under the heading of applied linguistics or Law. In any case, IR research is presently rather scattered, and teaching to undergraduates on this subject is fragmentary. There is presently little communication and collaboration between research groups. One notable exception is the CLEF (Cross-Language Evaluation Forum initiative by Hiemstra and Wessel Kraaij.

CLEF is the successor of the cross-language task which was organised for the past three years in TREC and it is aimed at Dutch/Flemish. This brings out another aspect: Holland is not an island, and the dutch language does not stop at the border. I have set out to bring together IR researchers from various disciplines and two countries into an interest group, under the name IRENE (Information Research In the Netherlands / Information Research In de Nederlanden).

On Tuesday, June 20 a group of about 20 researchers met at Nijmegen University to discuss the situation of IR in the Netherlands. Various groups briefly discussed their current research. I described the activities of [BCS - IRSG]. Floris Wiesman from Maastricht University volunteered to set up an IRENE website (still in embryonal state) with pointers to the various IR research groups. On 13 September 2000 Floris Wiesman and Jaap van de Herik from Maastricht University organized a full-fledged Information Retrieval workshop, attended by about 25 people. I then proposed to the SIKS directors to organise a two-day event in March 2001, an Advanced Course in Information Retrieval for PhD students from SIKS, aiming to fill the gap in their education. This proposal was warmly accepted by SIKS, so now I have to show deeds instead of words. I hope that my attempt to rally the forces of IR in the Netherlands will in some time provide [BCS - IRSG] with a robust partner in the Netherlands.

Nijmegen, Oct20, 2000.

INFORMATION RETRIEVAL RESEARCH AT HELSINKI UNIVERSITY OF TECHNOLOGY

**Mark Girolami
University of Paisley**

The primary focus of research at the Computer and Information Science Laboratory, University of Helsinki, is the theoretical development of computing paradigms that are inspired from neurobiology, arguably the most well known being the Self-Organising Map (SOM) proposed by Teuvo Kohonen. Despite the somewhat different motivations of the research from traditional Information Retrieval research, the laboratory has two current projects, which seeks to exploit the SOM within IR. The question remains open as to the impact; significance and ultimate utility that the output of such research will have on the wider IR community, nevertheless investigation of such models of adaptation have provided novel insights into the IR problem.

The two projects, which utilise the SOM model, are the WEBSOM and the PICSOM, WEBSOM concentrating on the organisation of text documents to possibly aid navigation through a large corpus and PICSOM focusing on content-based image retrieval.

WEBSOM is a method for organizing miscellaneous text documents onto meaningful maps for exploration and search. WEBSOM is based on the SOM (Self-Organizing Map) algorithm that automatically organizes the documents onto a two-dimensional grid so that related documents appear close to each other. The mathematical preliminaries, background, basic ideas, implications, and numerous applications of Self-Organizing Maps are described in the recent book Kohonen: Self-Organizing Maps (Springer, 1995; second extended edition 1997). The emerging field of text mining applies methods from data mining and exploratory data analysis to analyzing text collections and to conveying information to the user in an intuitive manner. Visual, map-like displays provide a powerful and fast medium for portraying information about large collections of text. Relationships between text items and collections, such as simi-

larity, clusters, gaps and outliers can be communicated naturally using spatial relationships, shading, and colors.

In the WEBSOM method the SOM algorithm is used to automatically organize very large and high-dimensional collections of text documents onto two-dimensional map displays. The map forms a document landscape where similar documents appear close to each other at points of the regular map grid. The landscape can be labeled with automatically identified descriptive words that convey properties of each area and also act as landmarks during exploration. With the help of an HTML-based interactive tool the ordered landscape can be used in browsing the document collection and in performing searches on the map.

An organized map offers an overview of an unknown document collection helping the user in familiarizing themselves with the domain however initial usability tests suggest that users may not favor such a model for navigation. Map displays that are already familiar can be used as visual frames of reference for conveying properties of unknown text items. Static, thematically arranged document landscapes provide meaningful backgrounds for dynamic visualizations of for example time-related properties of the data. Search results can be visualized in the context of related documents. Experiments on document collections of various sizes, text types, and languages show that the WEBSOM method is scalable and generally applicable. Preliminary results in a text retrieval experiment indicate that even when the additional value provided by the visualization is disregarded the document maps perform at least comparably.

PicSOM is a system for content-based image retrieval which is based on the SOM. The SOM is used to organize images into map units in a two-dimensional grid so that similar images are located near each other. PicSOM uses a tree-structured version of the SOM algorithm (Tree Structured Self-Organizing Map, TS-SOM) to create a hierarchical representation of the image database. During the queries, the TS-SOMs are used to retrieve images similar to a given set of reference images.

As a basis for retrieving images, the PicSOM system uses a combination of several types of statistical features, which are computed from the image content. Separate feature vectors can be formed to describe the colors, textures, and shapes found in the images. A distinct TS-SOM is constructed for each feature vector set and these maps are used in parallel to select the returned images. The image queries are performed through the World Wide Web and the queries are iteratively refined as the system exposes more images to the user.

PicSOM has been implemented as part of a project to study methods and systems for content-based image retrieval at Laboratory of Computer and Information Science at Helsinki University of Technology, Finland. Image retrieval with PicSOM is an iterative process utilizing the relevance feedback approach. The image query begins with an initial set of different images uniformly selected from the database. On subsequent rounds, the query focuses more accurately to the user's need. This is achieved as the system learns the user's preferences from the selections made on previous rounds.

Reference and Website

T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, V. Paatero, and A. Saarela. 2000 Self-Organization of a Massive Document Collection. *IEEE Transactions on Neural Networks, Special Issue on Neural Networks for Data Mining and Knowledge Discovery*, volume 11, number 3, pages 574-585. May 2000.

<http://www.cis.hut.fi/websom>

J. T. Laaksonen, J. M. Koskela, S. P. Laakso, and E. Oja. PicSOM - content-based image retrieval with self-organizing maps. *Pattern Recognition Letters*, 21(13-14): 1199-1207, November 2000.

<http://www.cis.hut.fi/picsom/>

Mark Girolami recently acted as the TEKES visiting professor in the Laboratory of Computing & Information Science, Helsinki University of Technology.

CALL FOR PAPERS

BCS-IRSG European Colloquium on IR Research

April 4 - 6, 2001, Darmstadt, Germany

<http://www.darmstadt.gmd.de/ECIR01/>

The colloquium on information retrieval research provides an opportunity for both new and established researchers to present papers describing work in progress or final results. These Colloquia were established by the BCS IRSG (British Computer Society Information Retrieval Specialist Group), and named the Annual Colloquium on Information Retrieval Research. Recently, the location of the colloquium has alternated between the United Kingdom and continental Europe. To reflect the growing European orientation of the event, the 23rd Colloquium in the series will be renamed The 23rd BCS European Annual Colloquium on Information Retrieval Research. This event is hosted by GMD, and co-sponsored by the BCS IRSG and GI. Relevant papers should address (at the theoretical, methodological, system or application level) the analysis, design or evaluation of functions like

- Indexing
- Information Extraction
- Data Mining
- Browsing
- Retrieval and Filtering
- User Interaction
- Composite documents
- Multimedia documents
- Hypermedia documents
- Active documents
- Distributed documents and databases
- Digital Libraries
- The Web

SUBMISSIONS

Authors are required to submit their paper, in English, by 12 January 2001. Papers should contain at most 7500 words and be double-spaced. The abstract should not exceed 100 words. The submission should include two copies of the paper in PDF or PostScript for-

mat: one full copy and one anonymous. Both files should be submitted by anonymous ftp to ls6.cs.uni-dortmund.de under the /pub/incoming/ ecir directory. The file names should reflect the title of the paper but the anonymous copy should have the prefix "anon". (To protect author privacy it will not be possible to list files in this directory). For the anonymous copy, the first page must contain the title of the paper and abstract, but no indication about the author(s) and affiliation(s). In addition, authors must send an email message to ecir@ls6.cs.uni-dortmund.de containing the title of the paper, the name of the file that has been submitted, the author name(s), and the author affiliation(s), plus complete contact information (mailing address, telephone, fax and e-mail) for the author to whom correspondence should be sent. Any queries regarding submission should be sent to:

goevert@ls6.cs.uni-dortmund.de

EXHIBITIONS

We invite interested companies active in document and knowledge management to show their products and latest developments. A special industrial session on products and applications - as well as experiences - will be considered. Any queries regarding exhibitions should be sent to: ulrich.thiel@gmd.de

CONTACTS

If you have any questions or problems concerning the colloquium, please contact: ulrich.thiel@gmd.de If you have any queries or problems concerning submitting a paper, please contact: goevert@ls6.cs.uni-dortmund.de

IMPORTANT DATES

Paper submission: **12 January 2001.**

Notification of acceptance: **20 February 2001.**

Final copy due: **15 March 2001.**

Conference: **4-6 April 2001**

PUBLICATION

All papers will be refereed. Following notification of acceptance, authors of selected papers will have until 10 March 2001 to make revisions in the light of

referees' comments. Papers will be published in the proceedings which will be circulated to all delegates during the Colloquium. Final papers will be published electronically (in some form to be decided). In addition, we expect to make arrangements for authors of good papers to be invited to make submissions to one of the major journals in the field.

FEES

Fees (not including accommodation) will be around EUR 180 (GBP 110), special rates for students, as well as for members of the participating organisations, will be available. In addition, CEPIS travel grants may be available for students.

COLLOQUIUM COMMITTEE

Ulrich Thiel, GMD-IPSI (Germany) (general chair)

Norbert Fuhr, University of Dortmund (Germany) (programme chair)

Christoph Baumgarten, Eurospider (Switzerland)

Micheline Beaulieu, University of Sheffield (UK)

Fabio Crestani, University of Strathclyde (UK)

Mark Dunlop, University of Strathclyde (UK)

Hans-Peter Frei, Infocons (Switzerland)

Pia Borlund Jorgensen, Royal School of Librarianship (Denmark)

Mounia Lalmas, University of London (UK)

Monica Landoni, University of Strathclyde (UK)

Massimo Melucci, Universita di Padova (Italy)

Josiane Mothe, Universite Paul Sabatier TOULOUSE III (France)

Mark Sanderson, University of Sheffield (UK)

Fabrizio Sebastiani, IEI-CNR (Italy)

Alan Smeaton, Dublin City University (Ireland)

Theo P. van der Weide, University of Nijmegen (Netherlands)

Christa Womser-Hacker, University of Hildesheim (Germany)

Ir-talks mailing list Ir-talks@dcs.gla.ac.uk,

<http://mailhost.dcs.gla.ac.uk/mailman/listinfo/ir-talks>

CALL FOR PAPERS SIGIR 2001

Twenty-Fourth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval September 7 - 12, 2001

New Orleans, Louisiana, U.S.A.

<http://www.sigir2001.org>

SIGIR is the major international forum for the presentation of new research results and the demonstration of new systems and techniques in the broad field of information retrieval (IR). The Conference and Program Chairs invite all those working in areas related to IR to submit original research contributions, posters, and proposals for tutorials, workshops, and demonstrations of systems. All contributions must be submitted electronically. See the Conference web site for further details:

AREAS OF INTEREST

- Formal Models, Language Models, Search Strategies, Fusion/Combination
- Machine Learning for IR, Text Data Mining, Clustering, Text Categorization
- Cross-Lingual Retrieval, Multilingual Retrieval, Machine Translation for IR
- Topic Detection and Tracking, Content-Based Filtering, Collaborative
- Filtering, Agents
- Web IR, Citation and Link Analysis, XML and Metadata, Digital Libraries
- Video and Image Access, Audio and Speech Retrieval, Music Retrieval
- Text Representation and Indexing, Information Extraction, Lexical
- Acquisition, Natural Language Processing for IR
- Performance, Compression, Scalability, Architectures, Distributed Search
- Interfaces, Visualization, Interactive IR, User Models
- Summarization, Question Answering

- Evaluation, Building Test Collections, User Studies, Experimental
- Design and Metrics

IMPORTANT DATES

January 29: Paper submissions due

February 28: Proposals for tutorials, workshops and posters due

April 20: Notification of acceptance for all submissions

May 25: Final camera-ready copy due

PAPER SUBMISSION GUIDELINES

Authors are invited to submit research papers, not exceeding 5000 words representing original, previously unpublished work, on or before January 29, 2001. Papers must be submitted electronically, via the submission web page. For requirements and details on submission of posters and on proposals for tutorials, workshops, and demonstrations of systems, please see the Conference web site at <http://www.sigir2001.org>.

CALL FOR PAPERS

The First Asia-Pacific Conference on Web Intelligence (WI-2001)

Maebashi TERRSA, Maebashi City, Japan, October 23-26, 2001

<http://kis.maebashi-it.ac.jp/wi01>

Paper Submission Deadline: **March 20, 2001**

The Asia-Pacific Conference on Web Intelligence (WI) is an international forum for researchers and practitioners (1) to present the state-of-the-art in the development of Web intelligence; (2) to examine performance characteristics of various approaches in Web-based intelligent information technology; (3) to cross-fertilize ideas on the development of Web-based intelligent information systems among different domains.

TOPICS

The following are some of the topics which are mainly related to IR. For

detailed topic list you may refer to the conference site.

- Web Information Retrieval:
- Approximate Retrieval
- Conceptual Information Extraction
- Image Retrieval
- Multi-Linguistic Information Retrieval
- Multimedia Retrieval
- New Retrieval Models
- Ontology-Based Information Retrieval
- Automatic Web Content Cataloging and Indexing
- Web Agents
- Information Filtering
- Recommender Systems

Please send suggestions and inquiries regarding WI-2001 to:rof. Ning Zhong (WI-2001), Department of Information Engineering, Maebashi Institute of Technology, 460-1, Kamisadori-Cho, Maebashi-City, 371-0816, Japan TEL&FAX: +81-27-265-7366

E-mail: zhong@maebashi-it.ac.jp

CALL FOR PAPERS JC DL 2001

Joint Conference on Digital Libraries

June 24-28, 2001

Hotel Roanoke, Roanoke, VA, USA

<http://www.jcdl.org/>

The Joint Conference on Digital Libraries is a major international forum focusing on digital libraries and associated technical, practical, and social issues. JC DL 2001 enhances the tradition of conference excellence already established by the ACM and IEEE-CS by combining the events that these professional societies have sponsored on an annual basis, the ACM Digital Libraries Conferences and the IEEE-CS Advances in Digital Libraries Conferences. Following JDCL will be an NSF PI meeting for the US Digital Libraries Initiative.

JC DL encompasses the many meanings of the term "digital libraries", including (but not limited to) new forms of information institutions; operational information systems with all manner of, digital content; new means of selecting, collecting, organising, and distributing digital content; and theoretical models of information media, including document genres and electronic publishing. Digital libraries are distinguished from information retrieval systems because they include more types of media, provide additional functionality and services, and include other stages of the information life cycle, from creation through use. Digital libraries can be viewed as a new form of information institution or as an extension of the services libraries currently provide.

The intended community for this conference includes those interested in aspects of digital libraries such as infrastructure; institutions; metadata; content; services; digital preservation; system design; implementation; interface design; human-computer interaction; evaluation of performance; evaluation of usability; collection development; intellectual property; privacy; electronic publishing; document genres; multimedia; social, institutional, and policy issues; user communities; and associated theoretical topics. Participation is sought from all parts of the world and from the full range of disciplines and professions involved in digital library research and practice, including computer science, information science, librarianship, archival science and practice, museum studies and practice, technology, medicine, social sciences, and humanities. All domains---academe, government, industry, and others---are encouraged to participate as presenters or attendees.

DEADLINES

January 9, 2001 Full papers, panel and tutorial proposals due.

February 7 Short papers, posters, proposals for workshops and demonstrations due

March 31 Final submissions due conference Web site <http://www.jcdl.org/>

Program Chair: Edward Fox and Christine Borgman, Dept. of Computer Science, Dept. of Information Studies, M/C 0106, Virginia Tech, UCLA, Los Ange-

les, Blacksburg, VA, 24061 USA CA,
90095 1520 USA Email:fox@vt.edu
Email:cborgman@ucla.edu

phone: +1 540 231 5113 phone: +1 310
825 6164

tel: +1 540 231 6075

fax: +1 310 206 4460

JOB OPPORTUNITIES IN INFORMATION RETRIEVAL AND DIGITAL LIBRARIES

Postdoctoral Research Assistant

Project Manager

The Department of Computer and Information Science of the UNIVERSITY STRATHCLYDE in GLASGOW are seeking to fill two R&D posts associated with an EU project funded under the IST Programme. Project MIND, which will focus on Resource Selection and Data Fusion for Multimedia Digital Libraries, is expected to start in January 2001 and last for 30 months.

The project is aimed at developing a variety of metadata for different generation methods media; designing algorithms for improving the selection of the most relevant collections of information; developing data fusion techniques for merging ranked lists of items retrieved from the selected collections; and evaluating these methods on different sources of data and cohorts of users. Other partners in the project are: University of Sheffield (UK), University of Dortmund (D), University of Firenze (I), and Carnegie Mellon University (USA).

At this stage we would like to invite those interested in applying for the above positions to contact us informally with an expression of interest and a short resume. More information involved about the project and the duties will be supplied upon request. Please contact either Forbes Gibb (forbes@dis.strath.ac.uk) or Fabio Crestani (fabioc@cs.strath.ac.uk) by email.

TWO RA POSTS AT SHEFFIELD UNIVERSITY IN INFORMATION RETRIEVAL

Two EU-funded projects are expected to start early in the New Year at the Information Studies department at Sheffield University and a research assistant (RA) post will result from each project. In anticipation of a formal job announcement early in January, potential candidates interested in the posts are requested to contact Mark Sanderson (m.sanderson@shef.ac.uk). The two projects are

Clarity (<http://clarity.shef.ac.uk/>). A 3-year EU funded project investigating interactive cross language information retrieval (CLIR). Sheffield's role is being run jointly between the Computer Science and Information Studies departments. The job of the Sheffield RA is to refine and user test a novel interface for CLIR. Candidates with a relevant post graduate degree with programming and HCI experience are desired: e.g. Java or C++ programming; experience in interface design; preferably multiple language skills particularly in Spanish or Finnish; information retrieval; and experience in user testing.

The Clarity consortium consists of 6 partners: Alma Media, BBC Monitoring, Bright Station (Martin Porter), Swedish Institute of Computer Science (Jussi Karlgren), University of Sheffield (Rob Gaizauskas and Mark Sanderson), and the University of Tampere (Kalervo Jarvelin).

MIND (<http://dis.shef.ac.uk/mark/mind/>). A 30-month EU funded project investigating the searching of heterogeneous databases. Sheffield's main task in the project will focus on the searching of spoken document collections with an emphasis on retrieval and interface design over actual speech recognition. In particular, searching collections held on remote databases is a focus of the project. Candidates with a relevant post graduate degree with programming and HCI experience are desired: Java or C++ programming; experience in interface design and user testing; information retrieval; and speech recognition.

The MIND consortium consists of 5 partners: Carnegie Mellon University

(Jamie Callan), University of Dortmund (Norbert Fuhr), University of Firenze (Alberto del Bimbo), University of Sheffield (Micheline Beaulieu & Mark Sanderson), University of Strathclyde (Fabio Crestani & Forbes Gibb)

The department of Information Studies at Sheffield has consistently gained the highest possible research rating and has a long tradition of high quality Information Retrieval research. It has strong links with the internationally renowned text-related research groups within the Computer Science department.

Mark Sanderson, Room 303 Tel : +44 (0) 114 22 22648, Department of Information Studies Fax : +44 (0) 114 27 80300, University of Sheffield, Western Bank mailto:m.sanderson@shef.ac.uk, Sheffield, S10 2TN, UK

<http://dis.shef.ac.uk/mark/>

PHD STUDENTSHIPS

Four Funded Research Studentships

School of Computer and Mathematical Sciences, The Robert Gordon University, Aberdeen

Owing to an expansion in our research activity, we have available one EPSRC Research Studentship and three University Research Studentships. You are invited to apply for research studentship for study leading to the award of MPhil/PhD.

The School has a reputation for high quality research in computing science, and we have strong, established research groups working in the fields of information retrieval, knowledge-based systems, and biomedical computing. Prospective PhD students applying for a research studentship are sought for the projects in the following areas: Biomedical Computing, Genetic Algorithms, Computer Graphics, Information Retrieval, Knowledge Based Systems.

All candidates should have a good Honours degree in computer/mathematical science or closely related discipline.

EPSRC Studentship

Candidates for the EPSRC Studentship should meet the EPSRC criteria for eligibility (see www.epsrc.ac.uk). Annual stipend is £8800; tuition fees are also covered. Applications are welcomed from both UK and EU nationals (although EU nationals may only qualify for tuition fees).

Three University Studentships

University Research Studentships are open to all. Annual stipend: £6800, and paid teaching up to £2000 per annum may be available. Tuition fees will be covered at the Home/EU rate, and overseas (i.e. non-EU) students will be expected to pay the difference between the Overseas rate and the Home/EU rate. Applications for University Studentships are welcome at any time. The University will provide an excellent research environment, and each student is provided with a high-spec workstation/PC. Research students are funded to attend conferences and workshops, and additional training courses for research students are provided.

Aberdeen is a superb place to live with good accommodation, excellent amenities and a large student population. The countryside surrounding Aberdeen provides myriad opportunities for hill walking, orienteering, canoeing, mountain biking, cycling, skiing and other outdoor pursuits. So, what are you waiting for? Don't delay - apply today!

For further details and an application form, please contact: Mrs Rhona Gibson, Research Support Officer, Academic Affairs Department, The Robert Gordon University, Schoolhill, Aberdeen, AB10 1FR. Email research@rgu.ac.uk

See our webpages for additional details on PhD research and research projects <http://www.scms.rgu.ac.uk/vacancies/studentships.shtml>

PhD Studentships

University of Sunderland

Four PhD Studentships in School of Computing, Engineering & Technology

University of Sunderland. The School of Computing, Engineering & Technology draws together the disciplines of Computing, Mathematical Sciences, Engineering and Technology. The school has a strong and growing research profile with 6 EPSRC and more than 10 EU-funded projects and over 100 registered research students. As a result of its success the School is looking to recruit 4 students, to take up 4, three year, full-time, PhD studentships from around 22nd January 2001. There is one studentship in the area of Total Productive Maintenance, one studentship in Medical Informatics and there are two further studentships in the areas of: smart sensors for distributed control systems, video surveillance, subspace identification models, advanced databases, web metrics and object orientated testing, neural modelling of language and information retrieval. Applicants should inspect the web site <http://www.cet.sunderland.ac.uk/postgrad/projects.html> for details of the research projects against which we are seeking applications for these studentships.

Applicants will have at least a first class or good second class honours degree in a relevant area and at least IELTS6/TOEFL550 English language proficiency scores, if they are not first language English speakers. For informal discussions in advance of submitting an application, applicants may get in touch with the staff whose name appears on the website against each project. Informal enquiries about any of these studentships may be made to Dr. Chris Bowerman Tel (0191) 5153629, email chris.bowerman@sunderland.ac.uk.

The award for each studentship covers fees and a bursary of £6800 per annum for a maximum of three years. If you have the knowledge and abilities to match our requirements then please submit your full CV with a letter of application quoting "CET Studentships" to Valerie Cornell, Informatics Centre, School of Computing, Engineering & Technology, St Peter's Way, Sunderland, Tyne & Wear, SR6 0DD

WHO'S WHO

IRSG Committee

Contact List 2000- 2001

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

Chair

Ayse S Goker-Arslan

School of Computer and Mathematical Sciences

The Robert Gordon University

St Andrew Street

Aberdeen AB25 1HG UK

Email: asga@scms.rgu.ac.uk

Vice-Chair

John Lindsay

School of Information Systems

Kingston University

Kingston Upon Thames KT1 2EE UK

Email: lindsay@kingston.ac.uk

Secretary

Jane Reid

Department of Computer Science

Queen Mary and Westfield College University of London

Email: jane@dcs.qmw.ac.uk

Treasurer

Margaret Graham

Institute for Image Data Research

University of Northumbria at Newcastle

Email: margaret.graham@unn.ac.uk

Colloquium - IRSG2000

Stephen Robertson

Microsoft Research Ltd

St George House

1 Guildhall Street

Cambridge CB2 3NH

UK

Email: ser@microsoft.com

Colloquium - IRSG2001

Ulrich Thiel

GMD-IPSI Dolivost. 15

D-64293 Darmstadt

Germany

Email: thiel@ darmstadt.gmd.de

Webmaster

Andrew MacFarlane

Open Muscat

St Johns Innovation Centre

Cowley Road

Cambridge CB4 0WB

Email: andym@open.muscat.com

Informer Editor

Mark Girolami
Department of Computing and Information Systems
University of Paisley
High Street
Paisley PA1 2BE UK
Email: m.girolami@paisley.ac.uk

Informer Associate Editor

Ali Asghar Shiri
Department of Information Science
University of Strathclyde
26 Richmond Street
Livingstone Tower
Glasgow G1 1XH UK
Email : shiri@dis.strath.ac.uk

General Members of the Committee

John Davies
Manager, KM Research
MLB3/65
BT Advanced Communications Technology Centre
Adastral Park
Ipswich IP5 3RE UK
Email: john.nj.davies@bt.com

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

Gareth Jones
Department of Computer Science
University of Exeter
Old Library
Email: G.J.F.Jones@exeter.ac.uk

Mounia Lalmas
Department of Computer Science
Queen Mary and Westfield College
University of London
Email: mounia.lalmas@dcs.qmw.ac.uk

Monica Landoni
Department of Information Science
University of Strathclyde
Livingstone Tower
26 Richmond Street
Glasgow G1 1XH UK
Email: monica@dis.strath.ac.uk

Kerry Rodden
Computer Laboratory
New Museums Site
Pembroke Street
Cambridge CB2 3QG UK
Email: kerry.rodde@cl.cam.ac.uk

Tony Rose
Retrieval Department Manager
Canon Research Centre Europe Ltd
Occam Road
Surrey Research Park
Guildford GU2 7YJ UK
Email: tgr@cre.canon.co.uk