InformeR

British Computer Society Information Retrieval Specialist Group

Greetings and Salutations

Welcome to the Winter 2001/2002 Edition of the **InformeR** newsletter, as you may have noticed there are a few changes to the look of the newsletter starting with the title and the logos. Aside from some of the structural changes, we have also added a feature area where we aim to present a focus article on information retrieval research groups throughout the UK. We'd also like to welcome a new member to the **InformeR** editorial team, Leif Azzopardi. He has recently started his PhD in the area of Information Retrieval at the University of Paisley.

This edition takes a look at the IR group at **The Robert Gordon University** in beautiful but blustery cold Aberdeen. We examine some of their goals, aims, recent developments and projects.

If you missed the **BCS-ECIR2001**, then read about some of the interesting presentations in a report by Alexi Vinokorov and get a taster of what this year's **BCS-ECIR2002** will be like. To find out when and where **BCS-ECIR2002** is on, check out the details in the **Up and Coming Events** section, and make sure you don't miss it!

We also present an article on the **Special Issue** that is coming out of the Journal of Intelligent Information Systems which features a collection of papers on Document Classification.

Are you packed and ready to go?

Of course we have the usual **Opportunities** section. Okay so we don't have any opportunities listed in this edition, but it doesn't mean there aren't any, we just couldn't find them − perhaps with a better information retrieval system…

And if you need to contact any of the BCS-IRSG members see the **Who's Who?** section.

We are always looking for interesting events and articles to fill the pages of the **InformeR**. So if you would like to notify other members of the group with up and coming events,

opportunities, book reviews, etc, then don't hesitate, send it in!

Happy New Year
Regards,
Editorial Team

Contents

IR Group Focus –
Robert Gordon University

It is hoped that each issue of **I**nforme**R** will present an article which focuses on IR research groups at particular institutes. This issue has a feature on the IR research group at the Robert Gordon University in Aberdeen.

The School of Computing at the Robert Gordon University, has a growing research reputation with significant research activity in a wide range of computing topics. The research targets technological advances that are directly relevant to industry and many of the projects have industrial collaborators in Aberdeen, the UK and worldwide.

Their computing research activities encompass the following areas:

- Knowledge Representation and Reasoning
- Knowledge Engineering Tools
- Genetic Algorithms
- Medical Informatics
- Virtual Environments
- Information Retrieval

In this article we take closer look at the Information Retrieval Group and some of their recent developments.

The members of their IR Group are:

Group leader
Prof. David J. Harper

Senior Lecturer
Dr. Ayse Goker-Arlsan

David recently received a SHEFC Grant for the SmartWeb project.

Ayse not only the chair of the BCS-IRSG but she also holds a black belt in karate!

Research Fellows
Dr. William Teahan
Dr. Daqing He

William has been applying compression-based language models to Information retrieval.

They also have several research students; Gheorghe Muresan, Christopher McGiffen and Jiusheng Yin.

Objectives

The IR group's focus is on devising new approaches to designing and building usable and effective IR Systems. Their work is primarily concentrated on four main areas:

interactive search environments, structuring and searching the World Wide Web, information mining and software architectures for constructing information retrieval systems and tools. This active research group has been built up over the past seven years and has been largely funded by industry groups such as the Union Bank of Switzerland IT Laboratory, British Telecom and the EPSRC. Recently the group has been working on two substantial projects; SmartWeb Technologies Centre funded by SHEFC and AmbieSense EU Project.

SmartWeb Technologies Centre

The University was awarded £748,000 by SHEFC Research Development Grant Programme to establish the SmartWeb Technologies Centre. This Centre is a new research facility at The Robert Gordon University and brings together university researchers and members of the computing industry and targeted sectors to collaborate on technically challenging projects. It aims to make "smart" internet products!

The motivation behind such an aim is that most current Web applications do not fully exploit intelligent computing technologies. However, users will increasingly demand tailored and personalised information services, while the information providers will want to automate these complex decision-making processes. Integrating intelligent technologies in Web applications is central to the development of new smart Web products and services, and the resulting benefits are potentially huge. For example:

- Smart interfaces will be personalised and will adapt depending on the user and the task.
- Smart websites will organise information in a flexible way,

and provide novel ways to access this information.

- Smart applications will enable better decision-making by users and allow them to collaborate more effectively on shared projects and activities.

The work of SmartWeb will focus initially on four important sectors.

- **e-Commerce:** smart applications and customisation will drive the massive growth that is forecast.
- **Healthcare:** remote monitoring and personalised information services will assist decision making for patients and health professionals.
- **Learning:** better ways of organising and delivering information will enhance distance learning and mobile working.
- **Government:** customised access to government information and decision making will improve public participation in the democratic process.

The SmartWeb Centre brings together expertise in Web Retrieval, Machine Learning and Medical Informatics from research groups led by Professor David Harper, Professor Susan Craw and Dr John McCall at The Robert Gordon University. It hopes to develop dynamic and sustainable partnerships with industry, and share in creating intelligent Web products and services that are relevant and applicable. In this way, the Centre will make a significant contribution towards achieving society's long

term goal of improved access to information services.

AmbieSense EU Project

As part of an EU Consortium, a group of IR researchers at RGU led by Dr. Ayse Goker-Arlsan, was involved in the successful negotiation of a 5 million Euro budget for the AmbieSense EU Project.

The project aims at developing intelligent agents that are mobile and distributed. Multi-agency and context-sensitive technology is a keyword. These agents will help users to get relevant information in the right context. The application domain is ambient, personalised, and context-sensitive information services in the travelling and tourism domain. The ambient technology will be tested and evaluated with real mobile users at an international airport, and in a large European city centre. Tourist information, infotainment and XML-based electronic maps will be put together on mobile and wireless computers the users carry with them whilst on the move. The users are travelling businesspersons, and tourists. Project Consortium Partners include SINTEF Telecom and Informatics, NTNU (Norwegian University of Science and Technology), CognIT, Reuters and other key content and software providers in the tourism industry.

Past and Current Projects At RGU

Here are some of the other interesting projects that the IR Group at Robert Gordon University, have been working on.
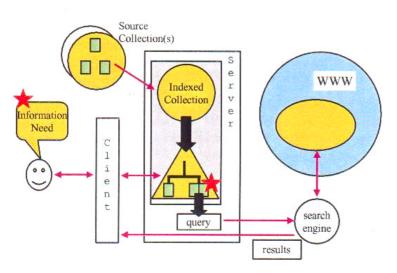

You've just got to have a cool logo!

Mediated Access to the Web

The WebCluster system enables structured, subject-specific portals to the World-Wide Web. WebCluster supports a new approach to searching called "mediated access", in which access to the web using search engines is mediated through a WebCluster portal. The subject-specific portals provide a context for searching, and assist the (novice) user in understanding the subject area and in formulating effective queries [8,6,9].

With WebCluster, a collection of documents converting a given subject (source collection) is automatically indexed, and the documents are clustered based on document-document similarity. The result is a hierarchical structure over the documents, which captures the topics covered in a subject area c.f. Yahoo! Categories. The WebCluster client supports a graphical interface, through which an end user explores the hierarchy using a combination of searching and browsing and so identifies a topic. Based on the topic, the WebCluster server generates a query, submits it to a search engine (or engines), and returns the resultant webpages via the client to the user. Essentially, the hierarchical structure acts as a subject-specific filter over the web. WebCluster has several possible applications. A collection of documents covering a particular domain, e.g. nutrition, allows a structured content portal to be built that provides more effective access to both the original source collection, and to nutrition webpages. Existing information resources with a hierarchical structure, (e.g. large reports with many sections, structured websites or intranets, and structured electronic reference works) can be loaded by WebCluster, and access mediated through the structure. Thus, access can be mediated from

The WebSmart System

any structured collection to any other searchable collection.

User Context Learning for Web Applications

Commercially available IR systems deal with searches in a vacuum; they do not consider the context of the information need nor the user's previous search patterns. They are building intelligent Web-based IR systems based on user modeling and machine learning techniques applied to real-life user data. Web user queries are analysed to detect patterns which can be exploited by incremental learning algorithms [5, 7]. In addition to publicly available data, a collaboration with Reuters Ltd. Provides intranet user data to enhance research. Although the amount and variety of information has been ever-increasing, the importance of user-adaptive IR systems has only been recognized since the wider usage of the Web.

This work builds on a User Context Learner [3] for traditional bibliographic IR system Okapi. Whilst there are similarities with information search/access in a library database environment, the Web does pose different challenges. A wider variety of users perform a wider variety of tasks. Also, the information available is dynamic and not necessarily vetted by an

authority. Our research recognizes such differences in order to handle Web users' queries more effectively.
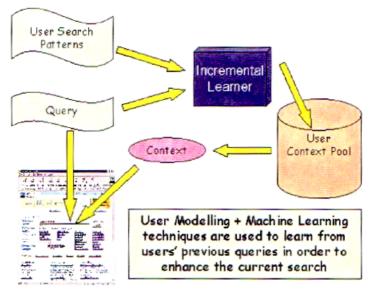
Language Modelling and Text Mining

Language modeling techniques based on Prediction by Partial Matching (PPM) [1] are being applied successfully to many natural language processing tasks (e.g. language and dialect identification, text compression, cryptography, text categorization, authorship attribution) [10] an various applications for automatically correcting words in text

(e.g. OCR and Chinese word segmentation [11]).

Text mining is the process of "(semi-)automated knowledge acquisition from linguistic resources". The standard approach applies manually designed tokenizers and grammars for the particular data being extracted but our approach uses supervised training and PPM models to detect sub-languages of text instead of explicit programming [12]. Language modeling techniques extract meaningful low-level information about the location of semantic tokens such as names, email addresses, locations, URLs and dates.

A Text Mining Toolkit has been developed for modeling sequential text [2]. It greatly simplifies the design of applications where the use of textual models is needed (e.g. data compression, text correction or segmentation) and will form the basis of further work in information retrieval, information extraction, question answering, text translation and text summarization.

The past and current projects were taken from the research portfolio of the IR research group and the group kindly permitted us to reprint them..

User Context Learning for Web Applications

Selected Publications

[1] J. G. Cleary and W. J. Teahan. Unbounded length contexts for PPM. *Computer Journal*, 40(2/3):67-75, 1997.

[2] J. G. Cleary and W. J. Teahan. An open interface for probabilistic models of text. In *Proceedings of the IEEE Data Compression Conference*, 1999.

[3] A. Goker. Context learning in Okapi. *Journal of Documentation*, 53(1):80-83, 1997.

[4] A. Goker. Capturing information need by learning user context. In *Proceedings of the Learning About Users Workshop at IJCAI'99*, pg21-27, Stockholm, Sweden, 1999.

[5] A. Goker and D. He. Analysing Web Search logs to determine session boundaries for user-oriented learning. *In International Conference on Adaptive Hypermedia and Adaptive Web-based Systems (AH2000)*, pg319-322, Trento, Italy, 2000. Springer-Verlag.

[6] D. J. Harper, M. Mechkour, and G. Muresan. Document clustering for mediated information access. In *Proceedings of the 21st Annual BCS-IRSG Colloquium on IR Methods, BCS Electronic Workshops in Computing Series*, Glasgow, Scotland, 1999.

[7] D. He. And A. Goker. Detecting session boundaries from web user logs. In *Proceedings of the 22nd Annual Colloquium of IR Research (IRSG2000),* pg57-66, Cambridge, England, 2000.

[8] M. Mechkour, D. J. Harper and G. Muresan. The WebCluster project: Using clustering for mediating access to the World Wide Web. In *Proceedings of the 21st Annual International SIGIR Conference on Research and Development in IR*, Melbourne, Australia, 1998. ACM Press.

[9] G. Muresan, D. J. Harper, A. Goker, and P. Lowit. ClusterBook, a tool for dual information access. In *Proceedings of the 23rd Annual International SIGIR Conference of Research and Development in IR*, 2000.

[10] W. J. Teehan. Text classification and segmentation using minimum cross entropy. In *Proceedings of the International Conference on Content-based Multimedia Information Access (RIAO 2000)*, pg943-961, Paris, France, 2000.

[11] W. J. Teehan, Y. Wen, R. McNab, and I. H. Witten. A compression-based algorithm for Chinese word segmentation. *Computational Linguistics*, In Press.

[12] I. H. Witten, Z. Bray, M. Mahoui, and W. J. Teehan. Using language models for generic entity extraction. In *ICML-99 Workshop on machine learning in text data analysis*, Stockholm, Sweden, 1999.

Further Information

For Further information about the IR group at the Robert Gordon University contact:
Professor David Harper
Email: djh@scms.rgu.ac.uk
School of Computing
St. Andrews Street
Aberdeen, AB25 1HG, UK

Or visit the group's website:
www.scms.rgu.ac.uk/research/ir

INFORMATION RETRIEVAL
SPECIALIST GROUP

BCS
THE BRITISH COMPUTER SOCIETY

Remember to check the details to BCS-ECIR 2002 in the Up and Coming Events Section.

The 23rd BSC-IRSG European colloquium on information retrieval research

The 23rd Annual BCS-IRSG Colloquium on Information Retrieval gathered researchers from more than 10 countries and covered a wide spectrum of topics including Support Vector Machines, pseudo-relevance feedback retrieval, time-restricted retrieval, text summarisation, text data mining, Hidden Markov Models and information retrieval from video content.

Specifically, the following list of research papers was presented:

- Flexible pseudo-relevance feedback via direct mapping and categorisation of research requests
- Applying heuristics to improve a genetic query optimisation process in information retrieval
- Document classification employing Fisher kernel derived from probabilistic hierarchic corpus representations
- A study on use of summaries and summary-based query expansion for a question-answering task
- Self-supervised learning for automatic text summarisation by text-span extraction
- From bags of words to bets on margins: text classification with support vector machines
- Information mining: use of the document dimensions to analyse interactively a document set
- Using data fusion to improve first story detection
- Searching and classifying the web using hyperlinks: a logical approach
- Categorising web documents in hierarchical catalogues
- A decision theoretic model for information retrieval supporting time constraints

- HMM-based passage models for document classification and ranking
- Retrieving scanned documents from a mixed-media document collection
- The power of word clusters for text classification
- Information extraction from video streams for retrieval by content
- Towards a domain analysis methodology for collaborative filtering
- Document filtering as an adaptive and temporally-dependant process
- Chatterbots and intelligent information search
- Evaluations of retrieval effectiveness in interaction information retrieval

Pseudo-Relevance Feedback

Sakai, Robertson and Walker presented an extensive range of distance measures for the Pseudo-Relevance Feedback technique on TREC6-8. They considered two cases direct mapping (request-to-request) and request categorization (request-to-group of requests). Their results suggested that the *offer weight* may be useful for determining the number of expansions terms for each given request and that request categorization may be more reliable than direct mapping. However, none of the techniques significantly outperformed classical pseudo-relevance feedback.

Thorsten Joachims presented text classification using Support Vector Machines. Thorsten's presentation was very informative and contained a lot of interesting points on specificity of text data for SVM classification and how this specificity can be utilised for efficient approximation of the leave-one-out model selection method. Using Reuters data, the application of the SVM classification on text documents was impressive. It was shown that the application of Vapnik's theory on prediction error of classifiers yields an accurate approximation of leave-one-out.

Unfortunately the paper in the proceedings is quite short only about two pages and omits some of the more fascinating details discussed in the presentation.

Document Classification

An extension of the Hidden Markov Model for document classification and ranking is proposed by Denoyer and Zaragoza. They use a HMM of a standard multinomial model in such a way that sequenced alternating relevant and irrelevant parts within documents are captured. Appropriately named the Irrelevant-Relevant-Irrelevant model (IRI). The IRI was compared against a baseline multinomial Naïve Bayes model using Reuters data and results showed a distinct improvement.

In another presentation by Slonim and Tishby from Hebrew University, Israel, the authors reported on a new word-clustering algorithm that enhances the performance of the Naïve Bayes.

Text Summarisation

An interesting investigation on the influence of text summarisation on interactive information retrieval was presented by the research group from University of Glasgow.
They suggested that text summarisation, which they defined as the selection of a subset of document sentences, which are representative of the document's content, may provide a useful source of evidence for question and answering tasks or for relevance feedback. During the presentation, the authors purported that the lack of use of relevance feedback is disappointing and that encouraging users to interact with IR system is a fundamental problem.

Another work on text summarisation was by Amini and Gallinari from the University of Paris 6. They presented an application of an unsupervised learning technique based on simple linear classifiers for text summarisation. Tested on Reuters data they reported a clear performance increase.

Text and Video Data Mining

Text data mining has become an important task with the millions of digital documents available on the Internet and information stores such as digital libraries. A research group from Research Institute of Toulouse, France, presented their work where they employed a variety of data mining methods for visualisation of large text corpora. The documents were first summarised and then contingency tables built using data mining methods. For instance, a factorial analysis was built which determined the semantic relationships between documents. These tables were then used for 3-D visualisation of the corpora, subsequent navigation through it.

The amount of video content is also increasing rapidly and demands automatic intelligent processing too. The problem of retrieving information from video content seems to be even more challenging than text retrieval, but researchers from Italy have been attempting to extract information from video streams. The authors demonstrated a number of commercial applications with an impressive spectrum of operations on video streams. One of the challenges that confronted the authors was the automatic categorisation of news videos. A system presented was able to classify *newscasters*, *reports*, and *weather forecasts*. Another application automatically determined the type of sport (such as soccer, tennis, gymnasia tics, etc) by analysing the video stream. The approach undertaken by the researchers included original techniques for extraction of sport-characteristic features like playground, players and audience.

Concluding Remarks

The Colloquium was a great experience, I was delighted to attend and meet many fellow colleagues from all over the world with whom I had the opportunity to exchange many interesting ideas.

Attending the Colloquium gave me an excellent opportunity to keep abreast of the latest achievements in the Information Retrieval Community. The experience has also brought many new and fresh ideas that I hope to incorporate into my research. Many thanks to the John Campbell Trust for sponsoring my trip to the Colloquium.

Alexei Vinokourov
Dept. of Computer Science
Royal Holloway College
University of London

Email: alexei@cs.rhul.ac.uk

Grants Awarded

The following grants have been recently awarded to fellow IR researchers. (If we have missed you out let us know and we'll gladly include your project in the subsequent editions, just send in the details.)

Eurovision - cross language information retrieval of images

Dr. M. A. Sanderson from the department of Information Studies at the University of Sheffield has been awarded an EPSRC Project grant to the value of £59,439. The project is to be completed over the next year and a half.

Abstract

This research project will investigate an almost entirely unexplored and important area of text-based information retrieval: the cross language retrieval of images by their textual descriptions. Through the automatic translation of users' queries into the language of the image descriptions, it becomes possible for anyone to query an image library regardless of the language(s) they know. Unlike cross language retrieval of text documents, where examining the retrieved objects requires the ability to read another language, judging the relevance of retrieved images places no such requirement on the searcher, thereby making the system usable to a wide range of people. With greater usability comes increased interaction from searchers where any feedback on the relevance of retrieved objects may be used to improve subsequent retrievals. Although such a system has been discussed in the literature, it has never been implemented or researched, therefore, results stemming from this investigation are novel. The proposed project is supported by a UK newswire company who are keen to use this technology to release their picture archive to a range of new foreign markets.

Personalisation of web searches through ostension and summarisation

Dr. J. M. Jose from the department of Computer Science at the University of Glasgow has been awarded an EPSRC Project grant to the value of £59,726. The project is to be completed over a year beginning in February 2002.

Abstract

The aim of the proposed research is the development of effective techniques to interact with web search engines. This specific proposal is based on the use of automatic query-based summarisation techniques for web documents. We have demonstrated that summaries are a popular and effective means of assessing documents that our approach to summarisation is more useful than the existing descriptions given by web search engines, and that summarisation techniques can lead to more efficient searching by a user. The presence of document summaries can also lead to more interaction with the user. This increased interaction can be used to implicitly gather information that can improve a user's search, e.g. by suggesting new query terms, recommending new web pages from the list of unseen documents or personalisation of summaries. To fully realise this benefit it is necessary to develop more sophisticated methods of creating summaries, in particular summaries that are tailored to the user's query, and to develop methods for extracting information from the user's interaction with the summaries. These developments are the main focus of this proposal.

Up and Coming Events

24th BCS IRSG European Colloquium on IR Research

Glasgow, Scotland, UK

March 25-27, 2002

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 Colloquium was renamed "European Annual Colloquium on Information Retrieval Research" from 2001. The previous five colloquia have been held in Darmstadt (2001), Cambridge (2000), Glasgow (1999), Grenoble (1998), and Aberdeen (1997).

This Colloquia aims to attract papers that 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

for the following types of documents and databases:

- Monomedia documents (e.g. text, images, audio, voice, video)
- Composite documents
- Multimedia documents
- Hypermedia documents
- Active documents
- Distributed documents and databases
- Digital Libraries
- the Web

Keynote Speaker –

Richard K. Belew

Richard is a Professor of Computer Science at the University of California in San Diego. He has special research interests in machine learning and free-text information retrieval. His interest in Information Retrieval systems first began when writing his thesis on Adaptive Information Retrieval, which investigated the application of connectionist learning techniques to the Information Retrieval problem. He believes that it has proven to be a very appropriate marriage of technology with application. For more information about his research and career visit his website at www-cse.ucsd.edu/users/rik/.

Sponsors

Colloquium is jointly sponsored by the BCS-IRSG, CEPIS and Memex Ltd.

INFORMATION RETRIEVAL
SPECIALIST GROUP

BCS
THE BRITISH COMPUTER SOCIETY

Proud Sponsors of ECIR2002

List of Accepted Papers

The organisers of ECIR 2002 received a number of good quality submissions in answer to the call for papers with over 50 papers submitted. Each manuscript was scrutinized by three members of the program committee in a double blind review. A total of 23 papers were eventually selected for presentation at the colloquium. Examining the list of papers selected (below) the reader will note that in addition to contributions from European authors, there are several papers to be presented by researchers from the USA and Asian continent. Below is a list of the papers to be presented: (Clustered manually by some of the conference committee members, might we just add!)

Document Categorisation

- E. Gaussier, C. Goutte, K. Popat, and F. Chen. *A hierarchical model for clustering and categorising documents*. Xerox Research Center Europe, Meylan. France.

- C. Peter, and C.H.A. Koster. *Uncertainty-based Noise Reduction and Term selection in Text Categorization*. University of Nijmegen, Dept. Comp. Sci. Toernooiveld 1, The Netherlands.

- M. Kongovi, J. Carlos Guzmán, and V. Dasigi. *Text Categorization: An experiment using phrases*. Dept. of Computer Science, Southern Polytechnic State University, Marietta, USA.

Web Information Retrieval

- T. Tsikrika, and M. Lalmas. *Combining Web document representations in a Bayesian Inference Network model using Link & Content-based evidence*. Department of Computer Science, Queen Mary, University of London, Mile End Road, E1 4NS London, UK.

- S. J. Kim, and S. H. Lee. *An Improved Computation of the PageRank Algorithm*. School of Computing, Soongsil University, Seoul, Korea.

- G. Lepouras, C. Vassilakis, and G. R.S. Weir. *Serving Enhanced Hypermedia Information*. Dept. of Informatics and Telecommunications, University of Athens, Athens, Greece. Dept. of Computer and Information Sciences, University of Strathclyde, Glasgow, UK.

Soft Computing

- M. Boughanem (1) and L. Tamine (2). *A study on using genetic niching for query optimisation in document retrieval*. (1)IRIT SIG Universite de Toulouse III, 118 Route de Narbonne, 31062 Toulouse, France. (2)ISYCOM Universite de Toulouse II, 5 Allees A. Machado, 31058 Toulouse Cedex, France.

- R. Rajapakse, and M. Denham. *Concept Based Adaptive IR Modelling FCA-BAM combination for Concept Representation and Encoding*. Centre for Neural and Adaptive Systems, University of Plymouth, UK.

Interactive Information Retrieval Systems

- E. Sormunen. *A Retrospective Evaluation Method for Exact-Match and Best-Match Queries Applying and Interactive Query Performance Analyser*. University of Tampere, Dept. of Information Studies, Finland.

- Finn, N. Kushmerick, and B. Smyth. *Genre Classification and Domain Transfer for Information Filtering*. Smart Media Institute, Dept. of

Computer Science, University College Dublin, Ireland.

- **Information Retrieval Models**
L. M. de Campos, J. M. Fernandez-Luna, and J. F. Huete. *A Layered Bayesian Network Model for Document Retrieval.* Dpto. de Ciencias de la Computacion e Inteligencia Artificial, E.T.S.I. Informatica. Universidad de Granada, Spain. Dept. de Informatica, Escuela Politecnica Superior.Universidad de Jaen, Spain.

- G. Amati, and C. J. van Rijsbergen. Term frequency normalization via Pareto distributions. Fondazione Ugo Bordoni, Roma, Italy. Computing Science Department University of Glasgow, Glasgow, Scotland.

- V. Lavrenko. Optimal Mixture Models in Information Retrieval. Center for Intelligent Information Retrieval, Department of Computer Science, University of Massachusetts, Amherst, MA 01003. USA.

Multimedia Information Retrieval
- G. Duffing and M. Smail. Organising and searching partially indexed image databases. UMR 7530 LORIA, Campus Sciences BP 239, 54506 Vandoeuvre-Les-Nancy France.

- D. Heesch, and S. Ruefer. *Fusion Descriptors for Content-Based Sketch Retrieval - A Comparative Evaluation of Retrieval Performance*. Imperial College, Dept. of Computing, London, England.

- Thomas Sodring, Alan F. Smeaton. Evaluating a Melody Extraction Engine. School of Computer Applications, Dublin City University, Glasnevin, Dublin 9, Ireland.

Structured Document Retrieval
- J. Vegas, P. de la Fuente, and F. Crestani. *A Graphical User Interface for Structured Document Retrieval.* Dpto. Inform´atica Universidad de Valladolid, Spain. Dept. of Computer and Information Sciences. University of Strathclyde, Glasgow, Scotland.

- T. Roelleke, M. Lalmas and G. Kazai. I. Ruthven, and S. Quicker. The *Accessibility Dimension for Structured Document Retrieval*. Queen Mary, University of London, University of
- Strathclyde, University of Dortmund.

Cross Lingual Information Retrieval
- C. J.A. McEwan, I. Ounis and I. Ruthven. *Mining The Web For Bilingual Text*. Department of Computing Science, University of Glasgow, Scotland.

- Douglas W. Oard, and F. Ertunc. *Translation-Based Indexing for Cross-Language Information Retrieval.* College of Information Studies, Hornbake(South Wing), University of Maryland, College Park, MD 20742 USA.

Query Modification
- R. W. White, I. Ruthven, and J. M. Jose. *The use of implicit evidence for relevance feedback in web retrieval.* Dept. of Computing Science, Glasgow University, Strathclyde University, Glasgow.

- P. Vakkari. *Subject Knowledge, Source Of Terms, And Term Selection In Query Expansion. An analytical study*. Dept. of Information Studies, University of Tampere. FIN - 33014, Finland.

- Benammar, G. Hubert, J. Mothe. *Automatic profile reformulation using a local document set analysis.* IRIT/SIG Université

Paul Sabatier. Toulouse III, 118 Route de Narbonne, 31062 Toulouse cedex 4 France. Institut Universitaire de Formation des Maîtres, Toulouse.

Fees

Fees (not including accommodation) are €205 (£130) for members of the BCS and/or associated societies, and €235 (£150) for non-members. Also a student rate is available, which costs €125 (£80). However, proof of studentship is required to be presented at the registration desk. Also there are a number of travel grants available to students on a first come first served basis.

Further Details

For further details regarding the location and travel, the programme of events, the committee, etc visit the Colloquia's website at www.cs.strath.ac.uk/ECIR02/ .

CALL FOR PAPERS

The 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval

Tampere, Finland.

August 11-15, 2002

SIGIR2002 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.

SIGIR 2002 welcomes contributions related to any aspect of IR, but the major Areas of Interest are listed below**:**

- 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, Mobile Applications

- Interfaces, Visualization, Interactive IR, User Models, User Studies

- Summarization, Question Answering

- Evaluation, Building Test Collections, Experimental Design and Metrics

Important Dates:

January 28, 2002
Paper submissions due

February 28, 2002
Proposals for tutorials, workshops, demonstrations and posters due

April 19, 2002
Notification of acceptance for all submissions

May 24, 2002
Final camera-ready copy due

June 11, 2002
Registration dead-line, normal rate

August 11-15, 2002
Conference dates

CALL FOR PAPERS

Joint Conference on Digital Libraries 2002

Portland, Oregon, USA
July 14-18, 2002

The Joint Conference on Digital Libraries is a major international forum focusing on digital libraries and associated technical, practical, and social issues. JCDL enhances the tradition of conference excellence already established by the ACM and IEEE-CS by combining the annual 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. JCDL 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, organizing, 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 also 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 such aspects of digital libraries 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

The beautiful Portland city host to JDLC2002

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.

Sponsors

Jointly sponsored by Association for Computing Machinery (ACM), Special Interest Group on Information Retrieval (ACM SIGIR), Special Interest Group on Hypertext, Hypermedia and the Web (ACM SIGWEB), the Institute for Electrical and Electronic Engineers Computer Society (IEEE Computer Society) and the Technical Committee on Digital Libraries (TCDL) in cooperation with The American Society for Information Science & Technology (ASIST)

Important Dates

January 14, 2002
Full papers, panel and tutorial proposals due

February 11, 2002
Short papers, posters, proposals for workshops and demonstrations due

April 8, 2002
Final submissions due

July 14-18, 2002
Conference dates

Further Details

For information about the format of submissions, content and length visit the JDCL website www.jdcl2002.org.

Conference Committee

General Chair
William Hersh
Oregon Health & Science University

Treasurer
Gary Marchionini
University of North Carolina

Program Chair
Lynetta Sacherek
Oregon Health & Science University

Student Volunteers Chair
David Maier
Oregon Health & Science University

Workshops Chair
Richard Furuta
Texas A & M University

Webmaster
Marilyn Bushway
Oregon Health & Science University

Local Arrangements
Lynetta Sacherek
Oregon Health & Science University

Registration
Dolores Judkins
Oregon Health & Science University

Tutorials Chair
Paul Gorman,
Oregon Health & Science University

Panels Chair
Sally Howe
National Coordination Office for Information Technology

Networking Coordinator
Jere Retzer
Oregon Health & Science University

Demos Chair
Jon Herlocker
Oregon State University

NSDL Support
Brandon Muramatsu
University of California Berkeley

Posters Chair
Lois Delcambre
Oregon Health & Science University

Sponsoring and Exhibiting
Michael Nelson
NASA Langley Research Center

Publicity
Marianne Afifi
University of Southern California

A/V Coordinator
Warren Harrison
Portland State University

CALL FOR PAPERS

CIVR '2002 – European Conference on Image and Video Retrieval

London, UK

July 18-19, 2002

Image and video storage and retrieval continues to be 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, are still limited. The Challenge of Image Retrieval series of conferences was originally set up to bridge the gap between the different communities with an interest in image retrieval.

This conference, the fourth in the series, aims to provide a European-wide forum for the discussion of challenges in the fields of image and video retrieval. A unique feature of this conference is the high level of participation from practitioners. Applications papers and presentations suitable for a wide audience are therefore particularly welcome. Topics of interest include but are not limited to:

- Query models, paradigms and languages for image/video retrieval

- Content-based indexing, search and retrieval of images
- Feature extraction and representation
- Visual perception and image/video retrieval
- Image/video search and browsing on the Web
- Similarity measures between images/video
- Semantic retrieval of images and video
- Pattern recognition techniques for image classification and retrieval
- Evaluation of image and video retrieval systems
- Studies of information-seeking behaviour among image/video users
- HCI issues in image/video retrieval
- Database architectures for image/video retrieval
- High performance image/video indexing algorithms
- Novel image data management systems and applications
- Image data management for multimedia systems

Further areas of interest are security and rights management pertaining to visual data.

Sponsors

CIVR is sponsored by The Institute for Image Data Research from the University of Northumbria, The British Computer Society Information Retrieval Specialist Group, The British Machine Vision Association, The Institution of Electrical Engineers and The Leiden Institute of Advanced Computer Science

Keynote Speaker

David Forsyth

David is an Associate Professor in Computer Science at the University of California, Berkeley. He has recently co-authored a book called Computer Vision – A Modern Approach. The book is available

David enjoys sampling and finding naked people.

from his web site (www.cs.berkeley.edu/~daf), along with numerous papers, which include titles such as "The Joy of Sampling" and "Finding Naked People".

Important Dates

March 4, 2002
Submission of full paper

April 10, 2002
Notification of acceptance

May 1, 2002
Camera-ready full paper

July 18-19, 2002
Conference dates

Conference Committee

Organising Committee Chair
John Eakins
University of Northumbria

Practitioner Program Chair
Margaret Graham
University of Northumbria

Publicity Chair
Richard Harvey
University of East Anglia

Technical Program Chair
Michael Lew
University of Leiden

Local Chair
Chris Porter
Getty Images, London

General Committee Members
Paul Lewis
University of Southampton

Alan Smeaton
Dublin City University

Peter Enser
University of Brighton

CALL FOR PAPERS

Special Edition of the Journal of Uncertainty, Fuzziness and Knowledge-Based Systems

A special issue on the management of uncertainty and imprecision in multimedia information systems is to be published soon in the International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS). The issue was produced with the help of the guest editors: Mohand Boughanem from the Institut de Recherche en Informatique de Toulouse, Fabio Crestani from the University of Strathcylde and Gabriella Pasi from the Consiglio Nazionale delle Ricerche.

The aim of the special issue is to address the problems relating to information access. With the diffusion of the Internet and the consequent increase in the production and exchange of multimedia information demands the development of effective systems for information access. The main aim of these systems is to provide a fast and effective identification of the information relevant to specific user needs; this is a complex decision making task, pervaded with vagueness and uncertainty, due to the fact that it requires an interpretation of both the user's information needs and the stored information. A promising direction to make effective information access is to model the uncertainty and vagueness intrinsic in the process of identifying relevant information, and to make the systems adaptive,

i.e. able to learn the user's concept of relevance.

Submissions of papers on the use of Soft Computing techniques for modelling the uncertainty, vagueness, subjectivity, and learning in information access are requested. Specific topics of interest include, but are not limited to the following:

- Information retrieval
- Database management
- Multimedia indexing and retrieval
- Information Filtering
- Multilingual information access
- Collaborative filtering
- Text Categorization
- Digital Libraries
- Text Mining.

Where the information access process is modelled and/or managed using Soft Computing techniques, like for example:

- fuzzy logic
- probabilistic reasoning
- multi-valued logics
- chaotic computing
- belief networks
- machine learning
- rough sets
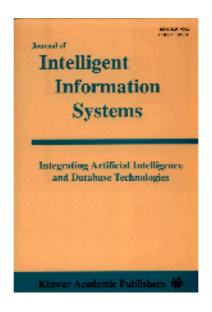- possibility theory
- neural networks

Further Details

For more details refer to the "Guidelines for contributors" included in each issue of the journal and as found on the journal website: http://ejournals.wspc.com.sg/journals/ijufks/mkt

Papers submitted to this Special Issue are subject to the usual IJUFKS peer review.

Important Dates

January 31, 2002
Submission of full paper

Special Topic Edition of the Journal of Intelligent Information Systems

The Journal of Intelligent Information Systems: Integrating Artifical Intelligence and Database Technologies has encouraged the presentation of research and development results focused on the integration of artificial intelligence and database technologies to create next generation information systems – Intelligent Information Systems.

These new information systems embody knowledge that allows them to exhibit intelligent behavior, cooperate with users and other systems in problem solving, discovery, access, retrieval and manipulation of a wide variety of multimedia data and knowledge, and reason under uncertainty.

The Journal of Intelligent Information Systems releases a Special Edition on Automated Text Categorisation in March, 2002. Fabrizio Sebastiani and Thorsten Joachims were guest editors of the special edition. Of the 22 papers submitted only 6 were selected for this issue and these are listed below.

T. Joachims, and F. Sebastiani
Guest Editors' Introduction to the Special Issue on Automated Text Categorization

A. Kabán, and M. Girolami
A Dynamic Probabilistic Model to Visualise Topic Evolution in Text Streams

N. Cristianini, J. Shawe-Taylor, and H. Lodhi
Latent Semantic Kernels

A. Vinokourov, and M. Girolami
A Probabilistic Framework for the Hierarchic Organisation and Classification of Document Collections

Chien Chin Chen, Meng Chang Chen, Yeali Sun
PVA: A Self-Adaptive Personal View Agent

P. Frasconi, G. Soda, and A. Vullo
Hidden Markov Models for Text Categorization in Multi-Page Documents

Y. Yang, S. Slattery, and R. Ghani
A Study of Approaches to Hypertext Categorization

Opportunities

If you would like to place an advertisement for an IR position in the newsletter please contact a member of the editorial team and we'll put it into the next edition.

Who's Who

IRSG Committee
Chair
Ayse S. Goker-Arslan
School of Computing
The Robert Gordon University
St Andrew Street
Aberdeen AB25 1HG UK
Email: asga@scms.rgu.ac.uk

Vice-Chair
Vacant.

Secretary
Jane Reid
Department of Computer Science
Queen Mary and Westfield College
University of London
Mile End Road
London E1 4NS UK
Email: jane@dcs.qmw.ac.uk

Treasurer
Margaret Graham
Institute for Image Data Research
University of Northumbria at
Newcastle
Newcastle upon Tyne NE1 8ST UK
Email: margaret.graham@unn.ac.uk

Webmaster
Andrew MacFarlane
Centre for Interactive Systems
Research
Department of Information Science
City University
Northampton Square
LONDON EC1V OHB UK
Email: andym@soi.city.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 Computing
The Robert Gordon University
St Andrew Street
Aberdeen AB25 1HG UK
Email: djh@scms.rgu.ac.uk

Gareth Jones
Department of Conputer Science
University of Exeter
Old Library
Prince of Wales Road
Exeter EX4 4PT UK
Email: G.J.F.Jones@exeter.ac.uk

Mounia Lalmas
Department of Computer Science
Queen Mary and Westfield College
University of London
Mile End Road
London E1 4NS UK
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

Tony Rose
Reuters Limited,
85 Fleet Street,
London EC4P 4AJ UK
Email: Tony.Rose@reuters.com

InformeR Team
Editor
Mark Girolami
School of ICT
University of Paisley
High Street
Paisley PA1 2BE UK
Email: m.girolami@paisley.ac.uk

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

Leif Azzopardi
School of ICT
University of Paisley
High Street
Paisley PA1 2BE UK