Informer

Number 21, Winter 2006/7
ISSN 0950-4974

BCS
INFORMATION RETRIEVAL

About Informer

Informer is the quarterly newsletter of the BCS Information Retrieval Specialist Group (IRSG). It is distributed free to all members. The IRSG is free to join via the BCS website (http://irsg.bcs.org/), which provides access to further IR articles, events and resources.

The British Computer Society (BCS) is the industry body for IT professionals. With members in over 100 countries around the world, the BCS is the leading professional and learned society in the field of computers and information systems.

Informer is best read in printed form. Please feel free to circulate this newsletter among your colleagues.

Welcome to the Winter Edition of Informer. It's around this time of year that I like to reflect a little on the events of the previous 12 months, and give some thought to plans for the next twelve. Uppermost in my mind (and that of a number of my colleagues) is our events diary – as you may recall, one of the highlights of 2006 was "Industry Day" - our venture into the new territory of extending our annual conference (ECIR) to include a day aimed exclusively at IR practitioners.

Now, to a degree all this should hardly be revolutionary – the BCS is a professional organisation after all, so practitioner events should be the norm. But the ECIR isn't like that, as its focus has traditionally been upon academic research, with a strong element of student participation. So, against that backdrop, Industry Day 2006 was indeed a new direction for us.

But what of Industry Day 2007? Well, this year is going to be a little different. One of the factors behind the success of ID 2006 was the choice of venue – in our London HQ we had a superb facility located in the heart of central London, which offered convenience both for one-day delegates and for those who'd just attended ECIR (a few miles away, at Imperial College).

But this year, ECIR is in Rome. Not surprisingly, therefore, we can't simultaneously use our London HQ and also be co-located with the main conference. So, what to do? Hold ID in Rome, or run a separate event in London? You'll be pleased, I hope, to learn that we've elected to pursue the latter option, and that ID 2007 will be held as a standalone event in London, on Weds May 23rd.

Moreover, this year's event looks set to be even bigger and better – we're still in the process of finalising the programme, but at the time of writing the lineup already includes a

number of well known names in the search business.

But there is one thing we have yet to resolve – the name. Evidently, we can no longer call it "Industry Day", since the concept of a day devoted to industry really only makes sense within the context of a larger, more academic event, such as ECIR. And since we're no longer co-located with ECIR, the name "Industry Day" seems, well, a little meaningless.

So, we need another name. In the absence of any better suggestions, how about "Search Solutions 2007"? OK, it's a bit contrived, but it alliterates nicely and does rather get straight to the point of what this event is all about.

Alternatively, if can you think of a better name, we'd love to hear from you. Similarly, if you'd like to get involved in this event, perhaps as a speaker, then just drop us a line at irsg@bcs.org. Meanwhile, mark it in your diary – **Search Solutions 2007, in London, on May 23rd**. The search practitioner event of 2007. Hope to see you there.

All the best,
Tony Rose
Informer Editor and Vice chair, IRSG
Email: irsg@bcs.org.uk

Feature Article:

There's a new KIDMM in town: A BCS knowledge community for information management

By Conrad Taylor

It has been suggested recently that it would be beneficial to create 'knowledge communities' within the British Computer Society around shared areas of concern, and that these communities might span several of the BCS Specialist Groups. The first of these communities will grow from an informal project of collaboration between more than a dozen BCS Specialist Groups that has sprung up in the last couple of years, called KIDMM.

KIDMM stands for 'Knowledge, Information, Data and Metadata Management', and is the working name of a grouping of about 45 individuals who share an email discussion list, a Web presence, and a developing programme of activities. The name is a bit of a mouthful, but expresses our core interest in developing a better common understanding about the 'I' part of IT; for the BCS, with its engineering background, can be argued as being stronger in its appreciation of Technology than of Information.

From catalogues to containers

When I'm asked to explain the rationale for KIDMM, I often turn the clock back 25 years and think about how little memory computers then contained: adequate for managing a simple database, but limiting if your profession involved handling information in textual form. In publishing (which is the background I come from), computers were being used for word processing in the 1980s, and for typesetting, but almost universally to produce a hard-copy, printed product.

There was already a class of people who were experts in managing these hard-copy knowledge and information assets rationally,

and we called them librarians. They had developed schemes for classification, such as the Dewey Decimal System. They also had a long tradition of collecting bibliographical data into catalogues, and they found that computer databases were useful tools not only for compiling catalogues, but also for machine-searchable records.

> **"The BCS, with its engineering background, can be argued as being stronger in its appreciation of Technology than of Information"**

The world today is very different. Computer memory is staggeringly huge, and networked globally: you could say that with the advent of the Web, the computer has swallowed the library! I certainly don't need to tell members of the Information Retrieval SG what this has meant in terms of the vast mass of data, information and knowledge assets – many of them hardly structured at all – with which we now have to deal.

Inverted-tree indexes, grep and Google are without doubt wonderful things, but other approaches are being called for that put content into context, for example tagging content with structured metadata, doing so with reference to controlled vocabularies, organising terms with ontologies, inter-relating these through Topic Maps, and so on; much of the effort aimed at making it possible for computers to be more efficient assistants in finding and acting on data and information in automated ways.

These have become issues for many communities. In the last ten years, the Electronic Publishing Specialist Group certainly found that publishers of all sorts, but especially scholarly and technical publishers, and people handling masses of text-based information in the museum and library communities, were having to get to grips with the problems of structuring, categorising and indexing their content. Meanwhile, other Specialist Groups also found themselves focusing on problems of managing data, or information, or knowledge, or all three; but we were all beavering away in our own corners, not talking to others much.

From metadata to meetings of minds

A couple of significant turning points came about in 2003. In January, the BCS Developing Countries Specialist Group ran a seminar, with John Lindsay as Chair, to see if we could influence British policy towards the UN World Summit on the Information Society. David Penfold and myself from EPSG participated in this meeting, which took as its focus 'Information Literacy'.

In April of the same year, the BCS Specialist Groups Assembly was held at Bletchley Park, and David Penfold was invited to give a talk about Metadata. This excited interest from a number of Specialist Groups, and in conversation it was thought to be a good idea to organise an inter-SG conference about metadata and its applications.

Though this conference didn't materialise, David was asked to speak about metadata at meetings of some other SGs, such as the Data Management SG, and to his initial puzzlement found that they had a completely different use of the term 'metadata' from that which is current in the world of publishing and librarianship! It became evident that it could be useful to try to bridge the multiplicity of understandings and expertises across the various IT specialisms and user communities which the BCS SGs represent.

> **Other SGs, such as the Data Management SG, had a completely different use of the term 'metadata'**

The breakthrough came at another SG Assembly, at an afternoon workshop convened to examine the relationship between the Forums which the BCS had created ('from on high', as it were), and the Specialist Groups which had formed from the bottom up. The point was made that the newly minted Health Informatics Forum was essentially the Health Informatics Committee, which had come together as a coalition of already well-founded Specialist Groups, with an obvious focus around computing and health.

So, the question was asked – were there any other shared areas of concern around which SGs might coalesce? Yes, said I – 'knowledge and information management'. A number of people agreed, and I circulated a notebook to collect email addresses. Interest was registered by individuals from about a dozen SG committees, and soon after we sent up an independent email discussion list on Topica to see where we might go from there.

EPSG decided to foster this collaboration, and I acted as the co-ordinator. Then on 6th March 2006 EPSG organised a discussion workshop for those who had expressed interest, and it is with the decision of those present in March to go forward from there that KIDMM can be said to have been 'born'.

You can get an idea of what KIDMM is about and who is involved from the Web site (see below). Also there is a comprehensive account of the discussions on 6th March 2006, and a collection of useful documents and essays.

Finding form and function

Now with an expanded and often lively email discussion list hosted on JISCmail, and a Web site hosted by EPSG, KIDMM's active partners are in the process of planning for a conference on 29th May 2007. This will be explicitly aimed at bringing diverse specialist communities within the BCS together, also drawing in other professional bodies who have contributions to make to our understanding of information management: the librarians and information scientists, the publishers and designers, for example.

> **"I'd love to see a couple of panels with content contributed by the IRSG"**

Another suggestion is to produce a parallel exhibition, ideally in the form of portable panels that can travel and be put up at different meetings and venues. Both the conference and the exhibition could be assembled in a modular fashion, based on contributed specialisms – I'd love to see a

couple of panels with content contributed by the Information Retrieval SG, of course!

As far as organisation is concerned, KIDMM has been invited to put up a proposal to the BCS Forums Board, with aims and objectives, a costed work plan and some form of governance acceptable to the BCS and, of course, the participating Specialist Groups and Forums. For KIDMM is not intended to replace the work of any part of the BCS; it will work best if it catalyses exchanges of knowledge between specialisms.

Note: IRSG members on the KIDMM email discussion are: Margaret Graham, Andy MacFarlane, Tony Rose.

Conrad Taylor is a writer, consultant and trainer in information design and electronic publishing. He has been active in the BCS Electronic Publishing Specialist Group for about 14 years and is currently its Chair. He can be contacted via: conrad@ideograf.demon.co.uk

Informer

BCS-IRSG Announcements from the Chair

by Leif Azzopardi

TLIR 2007 WORKSHOP: On the 10th of January, the BCS-IRSG ran the first International Workshop on Teaching and Learning Information Retrieval (TLIR2007) in London at the BCS-HQ. The aim of this workshop was to create a common space where IR lecturers and researchers can share their experiences and opinions in the field of IR teaching at different levels of educational. The presentations can be downloaded from the TLIR website.

DIR 2007 WORKSHOP: Recently, the BCS-IRSG has secured an "in cooperation" agreement with the Dutch Working Community on Information Sciences (WGI). Each year WGI hold the Dutch-Belgian Information Retrieval workshop. The primary aim of the DIR workshops is to provide an international meeting place where researchers from the domain of information retrieval and related disciplines, can exchange information and present new research development. The 7th Dutch-Belgian Information Retrieval Workshop (DIR2007) will be held in Leuven, Belgium on the 28th and 29th of March, the deadline for submissions is January the 10th.

SIGIR 2007: Just a quick reminder about the SIGIR 2007 deadlines. The full paper deadline for SIGIR 2007 is the 28th of January, 2007 and the poster deadline is the 25th of February, 2007.

ECIR 2007 CONFERENCE: The next European Conference in Information Retrieval will be held in Rome during April, 2007 (ECIR2007). On the 10th of December the ECIR PC was held to determine the programme for ECIR. They accepted 42 papers and 21 posters and this list of accepted papers is now available from the ECIR 2007 website. Congratulations to all submitting authors as the ECIR Chairs reported that the overall quality of papers was very high and that the

acceptance rate was less than 20%. With such competition it will always be difficult to get accepted, so to help improve your IR papers the BCS-IRSG has put together a guide on how to write good ECIR papers, which is available from the BCS-IRSG website.

ANNUAL GENERAL MEETING (AGM 2007): This year the AGM will not be held at ECIR 2007 in Rome. Instead, it will be held in London, at the BCS-HQ in May. We hope to co-located the meeting with the up and coming Industry Day and cordially invite all IRSG members to attend the Annual General Meeting. In the next edition of Informer, we shall publish the planned agenda of the meeting along with the date and time. If you have any issues you would like to raise at the next AGM then please feel free to contact me.

Finally, it is very important that we recruit new committee members each year and ensure that BCS-IRSG is moving forward and in new directions. If you would like to contribute to the IR community through the BCS-IRSG and become a more active IRSG member, then please contact myself or any other committee member about joining the committee.

Leif Azzopardi is an RCUK Research Fellow at the University of Glasgow, UK. His research interests include: formal models for information retrieval, distributed information retrieval and evaluation of information access systems. He can be contacted by email via: leif@dcs.gla.ac.uk

Book Review:

Information Extraction: Algorithms and Prospects in a Retrieval Context, by Marie-Francine Moens

Reviewed by Tom Betts

Information Extraction (IE) and Information Retrieval (IR) are core enabling technologies to the consumption of relevant content in our ever growing mass of unstructured data. In this text, Moens brings these two techniques together to illustrate how information derived using IE could be highly beneficial in IR systems.

Maybe you're thinking 'all these acronyms and buzzwords – what's going on?' Fortunately, the text is highly readable and aimed at both practitioners and researchers with some expertise in the textual domain. Emphasis is placed upon introducing IE, treating this as the emerging technology that we would like to integrate into existing retrieval systems. For example, we are introduced to typical tasks involved in IE such as the location and classification of named entities within a text, relationships between entities and events that involve these entities. The belief is that armed with this information we should be able to improve the performance of retrieval systems. One intuitive example is that IE may reduce the ambiguity of meaning present in words, which should help us better understand both queries and texts.

"the text is highly readable and aimed at both practitioners and researchers"

One trait that I offer particular praise to the author for is the pragmatic presentation of ideas. The introduction to IE takes us through the foundations of the discipline, right the way from symbolic rule-based systems to state-of-the-art techniques for pattern recognition using machine learning algorithms. The introduction to machine learning with text is comprehensive and a great reference point. We are also made to consider techniques for representing texts and the kinds of patterns we might want to automatically recognise for different tasks.

Recent advances in IE have used supervised learning techniques - relying heavily on exploiting a large set of (manually or otherwise) pre-classified solutions from which a classifier can be trained. Although we are taken through these techniques, the practical outlook provided by the author ensures that we are also moved swiftly onwards to consider unsupervised approaches to the same task. These techniques to not require large quantities of labelled data in order to train a classifier. Given that, as practitioners, we are rarely provided with sufficient pre-classified examples to use supervised approaches to classification it is particularly useful that we are introduced to unsupervised methods to IE.

"the text should be beneficial both to seasoned professionals and relative newcomers"

The final chapters of the book deal with the technicalities of combining IE with various retrieval models for IR systems. Further discussion of the benefits brought by integrating IE reiterate the aforementioned improvement in retrieval precision. In addition, 'softer' benefits are discussed, such as using IE to highlight facts within relevant documents, thereby easing the search process from the human perspective.

From a practitioners perspective, the concluding chapters are perhaps the most pertinent because the earlier material is put into perspective with a number of case studies. These example applications span legal, business and newswire texts, and include tasks such as intelligence gathering for law enforcement.

To take an overview of this book, the primary aim is to provide insight into the combination

of IE and IR techniques. In addition, it provides considerable contextual information about each of the respective technologies. The primary focus is on the textual domain, but applications of the technology are also considered for other domains. Given the clarity provided by the overview of these technologies, the text should be beneficial both to seasoned professionals in this area and relative newcomers.

Tom Betts is a data analytics consultant who specialises in unstructured data and the textual domain. His recent experience spans software development, consulting and research in both industry and academia. He can be contacted via: tom.betts@gmail.com

Search Solutions 2007

a.k.a
"Industry Day"

23rd May 2007
BCS HQ, London, UK

Following the success of Industry Day 2006, Search Solutions is a special day devoted to the interests and needs of IR practitioners. This forum presents an opportunity for commercial organisations and individuals to share their experiences, and for researchers to learn more about the issues and problems faced by IR practitioners in developing practical solutions for the information search and retrieval industry.

The programme is currently being finalised but at the time of going to press we have confirmed presentations from:

- Google
- FAST Search and Transfer
- BT
- Microsoft
- And many more!

To book your place at this unique event, keep an eye on the ECIR website - online registration will be available soon.

Workshop Report: The 3rd Asia Information Retrieval Symposium
by Mun-Kew Leong and Hwee Tou Ng

Introduction

Asia Information Retrieval Symposium (AIRS) 2006 was held at the Shangri-La Rasa Sentosa Resort in Singapore from 16-18 October 2006. The conference is the third in the AIRS annual conference series following on from the successful Information Retrieval with Asian Languages (IRAL) workshops which started in 1996. The first AIRS was held in Beijing, China, and the second AIRS was in Jeju, Korea. The AIRS series aims to bring together international researchers and developers to exchange new ideas and the latest results in IR. The scope of AIRS includes theory and practice of all aspects of both text and multimedia IR. Due to its roots in the IRAL series, there is also a strong focus on NLP-IR.

AIRS2006 received 148 submissions, the highest number since the conference series started. The submissions came from Asia, Australasia, Europe and North America. 34 submissions were accepted as regular papers (23%) and 24 as poster papers (16%). The co-organizing institutions were the A*STAR Institute for Infocomm Research, Singapore, and the Dept of Computer Science at the National University of Singapore.

Presentations

AIRS traditionally has two keynotes speeches. The first is academic and was presented by Prof Kui-Lam Kwok from City University, New York. His presentation was on "TREC" IR and Web IR: A Tale of Two Retrieval Environments, where he contrasted IR engines designed and evaluated in a TREC-style environment with those designed and evaluated in a Web-IR environment. Prof Kwok cited how the Cranfield collections were made obsolete by TREC, and explored the question of whether

TREC-style approaches would follow the same way with the strong momentum in Web-IR. With some discussion on the strengths and weaknesses of each approach, Prof Kwok concluded that they were complementary and could co-exist to help each other advance the field of IR.

The second keynote is normally from an industry speaker and is designed to make the participants consider the business aspects of search and retrieval. Our industry keynote speaker was Dr Daniel Lee, Yahoo!'s CTO for Asia, and also Convener of the ISO/IEC JPEG Standards Committee, who spoke on The Rise of Social Media and Opportunities for Research in Multimedia Search. Dr Lee gave an exciting multimedia presentation talking about the "long tail" phenomena and the rise of social media, i.e., user-generated media content. He explored the drive behind the business success of these internet models and the role that research and standards in IR can play in the necessary enhancements of these social phenomena.

The technical presentations were divided into 11 sessions which included the following: Text Retrieval, Search and Extraction, Text Classification, Clustering, IR Models, Web IR, CLIR, QA and Summarization, NLP-IR, Evaluation and Multimedia IR. The poster paper session was held in conjunction with the welcome reception on the rooftop of the Institute for Infocomm Research and provided participants with academic food for thought to accompany the dinner buffet of local (and spicy!) regional fare.

Finally, there was a special session on Medical Image Retrieval. This special session had two invited keynotes, one from Prof. Isabelle Magnin, the Director of CREATIS (Centre of Research and Applications in Signal and Image Processing), France on Medical Grids. The talk focused on various European IST grid projects and how they are used for biomedical services and applications including search and image processing. The second keynote was by Dr Patrick Brezillon from Laboratoire d'Informatique de Paris 6 (LIP6), France, who talked on Context-awareness and its application to medical image management.

Conclusions

The AIRS2006 conference was a success. There were over 90 participants from 10 countries, and included many PhD students. The full proceedings of the conference including both the oral and poster papers are published under the Springer imprint in the Lecture Notes in Computer Science (LNCS) series, Volume 4182, entitled Information Retrieval Technology. The next AIRS conference, in 2007, will be organized by the Harbin Institute of Technology (HIT) in China.

Dr. Leong Mun Kew is Principal Researcher at the Institute for Infocomm Research. His core research is in information retrieval, mobile and social information management, digital libraries and distributed multilingual search systems. He has 20 years of R&D and commercial experience including 2yrs as CTO in an IT startup. He can be contacted via: mkleong@i2r.a-star.edu.sg

Dr. Hwee Tou Ng is an Associate Professor of Computer Science at the National University of Singapore, Program Co-chair (Computer Science Program) of the Singapore-MIT Alliance, and a Senior Faculty Member at the NUS Graduate School for Integrative Sciences and Engineering. He received a PhD in Computer Science from the University of Texas at Austin, USA in 1992. He can be contacted via: nght@comp.nus.edu.sg

Workshop Report:

The First International Workshop on Teaching and Learning in Information Retrieval

by Michael P. Oakes

On Wednesday January 10th, the First International Workshop on Teaching and Learning was held at the BCS Headquarters in London. Although there were only 19 participants, it was a very informative and enjoyable meeting, and it is hoped that similar events will be held in future. The workshop proceedings are available at http://www.bcs.org/server.php?show=nav.00100v00500300100d001

Andy McFarlane introduced the workshop, pointing out that there has been no previous workshop on pedagogical research for IR. In addition, a lot of IR people in Universities have no experience in giving IR courses, their only opportunity for interacting with students being through project supervisions. We want to encourage more universities to teach this subject, so that students can use IR systems effectively, and/or conduct IR research, and to attract good people into our field.

Ana Cardoso-Cachopo described "IR base", an architecture for developing modular components for building IR systems. The components can communicate with each other, to provide a starting point for student projects, and provide teachers with examples for classes. The well tested, modifiable components allow experimentation with new ideas, since the experimenters do not have to code up from scratch. Components include an indexer, a text classifier, and optional probing interfaces which show what is going on inside a component - such as walking through a set of stemming rules for a query word, or showing how vectors are built in the vector model.

Stefanie Sieber described the design of e-learning courses at the University of Bamberg in Germany to support the teaching of IR. Using the DocBook tool, courses can be compiled from stored components. The tool supports Java applets which simulate such actions as pattern matching and clustering algorithms. As the simulations evolve, the corresponding Java code is simultaneously shown in another window, and moving red spots highlight the lines of code responsible for each step of the process. The tool also allows user feedback and the use of chat rooms. From student feedback it was found that the Java applets were less used than originally hoped, although the authors still feel these are potentially very useful.

Antonio Lopez-Herrera talked about teaching searching skills with fuzzy queries with a computer supported system. The interface shows the effect of adding and removing terms from the query on an updatable document-term matrix showing the weight of each term in each document. A graphical display shows which documents will be retrieved by various Boolean combinations of query terms, and the documents are colour coded according to their relevance. Juan Fernandez-Luna described the Garnata system which teaches probabilistic IR. Bayesian networks are represented graphically by qualitative (a ball and stick model of a Bayesian inference network) and quantitative (weight calculations for terms in documents) components.

> **"We want to encourage more universities to teach this subject"**

Andy McFarlane opened the second section by discussing the teaching of mathematics for IR, including set theory, Boolean models and term weighting. It is preferable for an IR expert to teach maths for IR (as opposed to a maths specialist), because they can use real IR examples to explain abstractions such as set theory. A tutorial style of delivery encourages active engagement by the students. At the end of the talk, Mark Sanderson suggested splitting up an IR course into a high level IR module (pitched at a "Discovery Channel" level) from

an optional hard maths module covering topics such as VSM. A third, case study driven, course could then be given in how to be a good searcher, for example using DIALOG.

To ensure that students have engaged in "deep learning", and understand the main principles of IR, Gareth Jones provides a problem-based assessment, where the students analyse a novel situation based on a research question. His students were asked to generate their own research proposal, on a theme such as "How would you summarise spoken documents?", and were asked to structure their proposals according to fixed templates.

Stefano Mizzaro asked when teaching web IR, should we cover the web first or IR first? Most classic IR books don't deal with the web, but Web IR covers additional topics such as web searching, crawling, link analysis, web technologies, languages and protocol, search with mobile devices, spam, dups, mirrors, and models of the web such as random, bow tie, and host graphs. There are some very good classic papers on these topics, which present new ideas without assuming prior knowledge. A show of hands at the workshop supported the teaching of classical IR before going on to web IR. Brian Mushens mentioned that in the old days of bibliographic databases it was difficult to put IR in the context of students' experience, but now we just have to mention "Google".

In the old days it was difficult to put IR in the context of students' experience, but now we just have to mention "Google"

David Bawden discussed curriculum development for IR, which forms part of the curriculum for Library and Information Science courses throughout Europe. Teaching IR needs the broader background of "information seeking", a broader set of things people do when searching for information including IR. This itself must be within a broader context of human information behaviour - why do people go to the information system, and what will

they do with the information after they have found it?.

In the final session, Eija Airio and Kai Haltunnen demonstrated the QPA (Query Performance Analyser) tool, which stores Finnish, Swedish and the TREC text collections, and contains retrieval software such as TRIP, InQuery, and Terrier. The tool has query manipulation resources, giving help with syntax formulation, can perform query translation for cross-language IR, and makes good use of the stored relevance judgements which come with the test collections to perform recall and precision calculations. QPA is thus able to display how any reformulation of a query or other change in the retrieval system affects performance. Earlier, Kai had discussed how a teacher, working with QPA can provide a "scaffold" for IR instruction. While the student formulates queries with QPA, at each step the teacher can provide hints such as suggesting parts of the query for modification, or comments such as "Why did this happen?"

Michael Oakes is a Senior Lecturer in the School of Computing and Technology at the University of Sunderland. He teaches modules in Information Retrieval to second and third year undergraduates, and a Master's module in Natural Language Processing. He can be contacted via:
Michael.Oakes@sunderland.ac.uk

Informer

Forthcoming Events

Edited By Andy MacFarlane

Search Engine Strategies
ExCel London, 13-15 February, 2007. The Premier Event for Search Engine Marketing & Optimization.
http://www.searchenginestrategies.com/sew/london07/index.html

4th Conference on Information Retrieval and Its Applications
Ecole des Mines de Saint Etienne (Loire - France) 28-30 March 2007. A French IR conference.
http://www.emse.fr/CORIA07/index.php?lang=en

New Directions and Challenges of Near-Duplicate Detection
Session at IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007), Honolulu, Hawaii, 1-5 April 2007. A session on detecting duplicates (an interesting IR problem) at a general AI conference.
http://ir.iit.edu/%7Ealek/dedup07.html

29th European Conference on Information Retrieval (ECIR 2007)
Rome, Italy, 2-4 April 2007. The IRSG's annual conference held at Fondazione Ugo Bordoni.
http://ecir2007.fub.it/

10th International Conference on Business Information Systems (BIS 2007)
Poznan, Poland, 25-27 April 2007. General information systems conference with a theme on IR and several workshops of interest.
http://bis.kie.ae.poznan.pl

Search Engine Meeting
Boston, USA, April 23-24, 2007. The Search Engine Meetings bring together commercial search engine developers, academics and corporate professionals to learn from each other.
http://www.infonortics.com/searchengines/

Special track on Context in AI Tools and Applications (CAITA 2007)
Key West, Florida, USA, 7-9 May 2007. Of interest to members who work in the area of IR and context.
http://www-poleia.lip6.fr/~jbrezillon/FLAIRS-2007/

Sixteenth International World Wide Web Conference (WWW 2007)
Banff, Alberta, Canada, 8-12 May 2007. The most highly regarded web conference, with a theme on IR. http://www2007.org/

NTCIR-6 Pre-Meeting Workshop (PMW) on Information Access Evaluation
National Institute of Informatics, Tokyo, Japan, 15th May 2007. An open session prior to the NTCIR evaluation, dealing with retrieval in Asian languages.
http://research.nii.ac.jp/ntcir/ntcir-ws6/pmw-en.html

The 2007 IEEE International Symposium on Data Mining and Information Retrieval (IEEE DMIR-07)
Sheraton Fallsview Hotel, Niagara Fall, Ontario, Canada, 21-23 May 2007. A conference on the issue of information retrieval and data mining.
http://www.cis.drexel.edu/faculty/thu/DMIR-07/DMIR07.htm

Search Solutions 2007
BCS HQ, London, 23 May 2007. The IRSG's premier practitioner event. http://irsg.bcs.org/

LIBRARIES IN THE DIGITAL AGE (LIDA) 2007
Dubrovnik and Mljet, Croatia, 28 May - 1 June 2007. A digital libraries conference.
http://www.ffos.hr/lida/

Large-Scale Semantic Access to Content (Text, Image, Video and Sound) RIAO 2007
Pittsburgh, USA, 30 May to 1 June 2007. A tri-annual conference on both scientific and industrial solutions to information processing problems.
http://www.riao.org/

The Eleventh International Conference on ARTIFICIAL INTELLIGENCE and LAW (ICAIL 2007)
Stanford University, Palo Alto, CA USA, 4-8 June 2007. An AI and Law conference which has one theme on information retrieval. http://iaail.org/

International Workshop on Pattern Recognition in Information Systems (PRIS-2007)
Funchal, Madeira, Portugal, 12-13 June 2007. A pattern recognition workshop with several IR themes, such as content image retrieval and multimedia categorization.
http://www.iceis.org/workshops/pris/pris2007-cfp.html

Fifth International Workshop on Content-Based Multimedia Indexing (CBMI-2007)
Bordeaux, France, 25-27 June 2007. A workshop on the indexing and retrieval of multimedia information, including extraction, evaluation etc.
http://cbmi07.labri.fr/

11th International Conference on User Modelling (UM2007)
Corfu, Greece, 25-29 June, 2007. A general user modelling conference, which will be of interested to members interested in modelling user behaviour in IR. http://www.iit.demokritos.gr/um2007/

Information: Interactions and Impact (i3)
The Robert Gordon University, Aberdeen, Scotland 25-28 June, 2007. A conference on the impact of user interactions in information, and the technology which mediates this – including IR.
http://www.i3conference.org.uk/

Summer School on Multimedia Semantics:
Analysis, Annotation, Retrieval and Applications Glasgow 2007, 15-21 July 2007. A summer school on Multimedia Semantics of interest to members who work on Multimedia IR.
http://www.dcs.gla.ac.uk/ssms07/

The 30th Annual International ACM SIGIR Conference (SIGIR 2007)
Amsterdam, Netherlands, 23-27 July 2007. The top IR conference. http://www.sigir2007.org/

5th IAPR International Conference on Machine Learning and Data Mining
Leipzig, Germany, 18-20 July 2007.
A general AI conference with several themes of interest to IRSG members (e.g. classification, image retrieval etc). http://www.mldm.de/frame.htm

Search Technology Trexy.com wins BCS Award

London, 8 December 2006. New search technology, Trexy.com, has been awarded a British Computer society (BCS) IT Professional award in the Services category.

The annual award encompasses excellence and innovation in computing within the context of business value and social benefit. The award was judged on originality including developments of novel approaches to systems, addressing social issues, the level of exploitability and the usage level.

Trexy's CEO and inventor, Nigel Hamilton accepted the award at the ceremony. "I am honoured to win the services award from the BCS for Trexy's technical novelty and inventiveness."

"I faced an ambitious technical challenge when taking a simple idea and applying it to Internet searching - universally. This award is great recognition of Trexy's unique approach to information retrieval which harnesses the collective effort of online searchers," he said.

Other nominees in the Services award category included: The Royal Bank of Scotland, Cosmic, and Alliance & Leicester PLC. The winners of the BCS IT Professional Awards were announced at the BCS Awards ceremony at the Grosvenor Hotel in London on the 7th of December 2006.

Contacts

Web: http://irsg.bcs.org/
Email: irsg@bcs.org.uk
Subscriptions: http://irsg.bcs.org/membership.php
ISSN: 0950-4974

To subscribe, unsubscribe, change email address or contact details please visit http://irsg.bcs.org/ or email irsgmembership@bcs.org.uk.

The IRSG is a specialist group of the British Computer Society.
To automatically receive your own copy of Informer, simply join the IRSG via the IRSG website.