

In This Issue

Editorial	1
<i>by Tony Rose</i>	
Conference Review	2
"ECIR 2006" <i>by Alex Bailey</i>	
Forthcoming Events	5
<i>Edited by Andy MacFarlane</i>	
Book Review	6
"Web Content Delivery" <i>Reviewed by Andrew Neill</i>	
Feature Article	8
"New Players in the Information Society" <i>by Theo Huibers</i>	
Product Update	10
"Trex.com: New Search Trails" <i>by Megan Hamilton</i>	
Feature Article	11
"A Tribute to Karen Sparck Jones" <i>by John Tait</i>	
Contact Information	13

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.



Well, they said it couldn't be done. But we did it anyway! As many of you will know by now, Industry Day 2006 took place on April 13, and, by all accounts, was a great success.

OK, I'm exaggerating a little - but there's no denying ID 2006 was something of a risk. With any new event you need to convince people of its significance - just because you can see its value doesn't mean that others will share your opinions. There's nothing more embarrassing than having your guest speakers present to a room full of empty seats. You have to articulate the appeal of your event to the right audience - and in that respect, we didn't do too badly. As you'll see from Alex Bailey's review of ECIR on page 2, ID 2006 was attended by around 50 people, and benefited from a line-up of speakers that featured presentations from major solution vendors such as Google, Microsoft, FAST Search and many more. The afternoon session also explored the way in which search technologies are changing the nature of business models in related industries (such as publishing), and in this respect I'm pleased to introduce Theo Huibers' as the author of our feature article on page 8.

But it wasn't all a success story. In hindsight, our marketing could have been much better - speaking as a "techie" myself, marketing and promotion doesn't exactly come as second nature, and looking back on our promotional plans we could have done a lot more to appeal to more mainstream information professionals within the digital libraries community and other end-user groups.

But overall, it was a good start, and definitely something to build on for next year (in fact, we have already identified May 16th as our provisional date for ID 2007). In the meantime, we'll be running an online survey to gather feedback on what you thought of ID 2006. For now, this survey will be aimed at just the ID attendees, but we'll be publishing

the results to everyone in the next issue of Informer.

More generally though, we'll also be soliciting feedback on other IRSG issues - notably Informer itself. To this end, we'll be running another short survey, this time to gather feedback about what you like and what you don't like about Informer, and how you think it could be improved. We'll be writing (via email) to each of you personally with further details later in the year. Of course, we appreciate that (like many professionals) IRSG members are busy people, so as an added incentive we'll be offering free IR textbooks for two lucky people whose names will be drawn at random from the first 50 responses we receive.

If you didn't get to attend ID 2007 in person, you can still access the slides, which we'll be making available from the event website (<http://ecir2006.soi.city.ac.uk/index.php?page=indust>). Inevitably, one or two of the talks will need a little editing before they can be made public, but we hope to have the full set available before too long. So in the meantime, enjoy the presentations, and don't forget to give us your thoughts on Informer.

Best regards,
Tony Rose
Informer Editor and Vice chair, IRSG
Email: irsg@bcs.org.uk

Conference Review: ECIR & Industry Day 2006

By Alex Bailey



This month saw the BCS IRSG's annual conference take place at Imperial College in London. This, the 28th European Conference on Information Retrieval (ECIR), was a record-breaking event in many ways. There were a record-breaking

number of paper submissions, a record-breaking number of papers presented, and a record-breaking number of attendees for the conference.

This marks the continuing rise in popularity of the event, which was first held in Leeds in 1979 as a colloquium focused on student work. It was then a forum to allow students to present work-in-progress and receive constructive feedback from established academics in the IR community. The event has now grown, reflecting the world-wide rise in exposure of the field, and in 2003 in Pisa it was renamed from a colloquium to a full-blown conference.

The conference still maintains a student-friendly atmosphere with 73% of accepted papers and 66% of accepted posters having a student as a first author. The attendance is largely dominated by students, but there is also a notably increasing presence of industry representatives filling the lecture theatres and coffee breaks.

The keynote speech was given by David Hawking of CSIRO, Australia. David gave an overview of the history of experimentation in IR including well-known experiments such as the Cranfield, MEDLARS and Lancaster studies. These were essentially conducted in real environments with real information needs from real users, using non binary relevance criteria such as highly relevant, minimally relevant not relevant. [TREC](#) has changed all this: the evaluation mechanism has become very popular, but the tasks are artificial and simulated using binary relevance criteria for

the most part. David pointed to some work done by Turpin and Hersh that questioned the validity of such an approach. It is clear from the web track experiments that all documents are not equally relevant, which reflects the type of search conducted in the enterprise. Enterprise search requires real time access control (an organisation will not want all employees to be able to find every single document e.g. for privacy reasons). TREC ad-hoc evaluation methods are therefore not appropriate in this environment. It is important to collect data such as click logs to understand user information searching behaviour - we therefore have to go back to the type of experiment conducted in the Cranfield, MEDLARS and Lancaster studies.

“It is important to collect data such as click logs to understand user information searching behaviour”

The rest of the presentations were split into sessions as usual, however to accommodate the increased number of presentations in the usual three days the conference included parallel sessions on the second day. Sessions covered the following familiar core topics: Formal Models, Design and Evaluation, Web Search, Cross-Language Retrieval, Topic Identification and News Retrieval, Clustering and Classification, Document & Query Representation and Text Understanding, Performance and Peer-to-peer Networks, Structure/XML, Multimedia, and Refinement and Feedback. There was also a new session with two papers devoted to Genomic IR. It is encouraging to see the conference broadening out to applications in Bioinformatics. We should not forget that IR is important for many of the Informatics disciplines.

One of the trends visible from the presentations is the increasing importance of XML and structured document retrieval with four papers devoted to the subject. As more documents are marked up using the accessible and popular XML format then this is certain to grow in interest. Also a paper on maintaining dynamic indexes – something very important for desktop search applications – was rightly awarded the best student paper prize for

Stefan Büttcher. The student poster prize went to Peter Wilkins for his poster on automatically generating feature weights for content-based image retrieval. Another notable new application was the use of clustering to improve the effectiveness of IR on a small-screened mobile device.

A new evening session this year was the ‘speaker’s corner’. This was an improvised poster session for anyone who wanted to take a colour marker to a flipchart and get on an imaginary soap-box to express their ideas. A representative from [Trex](#) took the opportunity for some informal marketing, Leif Azzopardi drew quite a crowd with an idea to log the progress of the conference with instant feedback via a wireless blog, and many others were seen discussing various ideas in every corner of the room.

The banquet was held in the ‘Conservatory’ at the top of the Millennium hotel just near Gloucester Road tube station. As far as conservatories go, this one was particularly welcoming, amply seating the near 200 delegates in a light and airy glass room on the roof of the hotel, complete with real palm trees (we checked). The food was good, the conversation was engaging, and the wine flowed nicely. No wine bottle was left unfinished thanks to the usual suspects from UCD, Imperial, and the IRSG committee.

“all three of the major web search companies are in a phase of recruitment”

It was interesting to note that all three of the major web search companies (Google, Yahoo, and Microsoft) are in a phase of recruitment as they fight for the top spot in today’s fierce market. On the first morning of the conference Google had a couple of representatives giving out freebies and openly discussing recruitment for their London and Zurich offices. Yahoo were seen having discreet discussions, no doubt for the new research lab in Barcelona, and Microsoft Research Cambridge are hiring, though if they were doing it at the conference it was very discreet. It must be a good time for finishing PhD students, though my advice

would be to make sure you know all the relevant data structures and algorithms!

And this year for the first time, the conference was followed by a special day dedicated to practitioners in information retrieval, knowledge management, and related industries. The BCS IRSG Industry Day was held at the impressive BCS London Office near Covent Garden and was attended by 50 delegates, roughly equally split between practitioners and academics. The day started with a couple of talks on tuning enterprise search, one from David Elworthy of Google, and the other from Steve Robertson of Microsoft! A lively discussion ensued during questions and it was interesting to see the two speakers side by side. Martin Porter then gave a detailed talk on how you might use term ranking to help users understand their search results.

After lunch, Tony Hart from FAST Search and Transfer gave a professional talk on knowledge management and business intelligence. Representatives from FAST also had a stand outside the room to give out freebies and demonstrate their software solutions. David Phillips from Corpora Software then explained the linguistic technology used in their suite of software products for knowledge management. Clive Flashman from the National Patient Safety Agency gave an entertaining presentation explaining the trials of creating a wide-ranging monitoring system that has to deal with the enormity of the NHS. The final presentations were from Theo Huibers of KPMG who gave an alternative and entertaining look at the changing landscape of content providers and consumers, and Wendy Johnson who showed how a collaborative web portal could be designed and built to target a particular industry sector, namely life sciences.

The event was a great success and the talks were lively and informative. It was good to see the delegates in lively discussion over lunch and in the break. We will be holding another event next year, so please contact me, the IRSG Event Co-ordinator if you are interested in speaking or helping out in any way.

It's great to see the conference expand and diversify, though with the unstoppable advance of information retrieval and the

increasing commercialisation of IR through web search, desktop search, and enterprise search, and all the associated technologies, has ECIR lost its student focus? Compared to the events of the past it has. However a larger conference brings the ability to introduce new features, such as the speaker's corner, where students can air their ideas in an open environment. It would also be very possible to include a doctoral consortium linked to future events, where students can receive feedback from a panel of experienced researchers.

As a committee member of the IRSG I will spare a few words for the committee. The reason I mention this here is that the IRSG is responsible for assessing bids for future ECIR conferences. So, if you are interested in shaping the conference or any other IRSG event, actively joining the committee is one way to do it. This year we will be holding an event to discuss the reviewing process for ECIR to ensure that we maintain the quality of the conference.

Next year the conference will take place in Rome from the 2nd-4th April.
<http://ecir2007.fub.it/>

Alex Bailey is experienced in information retrieval and natural language technologies underlying today's knowledge management products. He worked for 5 years at Canon Research Centre Europe investigating and developing document clustering, classification, summarisation, and information extraction technologies. He is now working as an independent consultant fostering links between industry and academia. Alex is also the One-Day Event Co-ordinator for the IRSG. He can be contacted via: alex.bailey@bcs.org.uk

Forthcoming Events

Edited By Andy MacFarlane

The Future of Web Search

Universitat Pompeu Fabra, Barcelona, 19-20 May 2006. A two day workshop on web search.
<http://grupoweb.upf.es/workshop/>

15th International World Wide Web Conference (WWW 2006)

Edinburgh, UK, 22nd –26th May 2006. A special conference track devoted to Web Search.
<http://www2006.org/tracks/search.php>

Libraries in the Digital Age (LIDA 2006)

Dubrovnik and Mljet, Croatia. 29 May - 4 June 2006. An annual course and conference on Digital Libraries. <http://www.ffos.hr/lida/>

1st International conference on scalable information systems (INFOSCALE 2006)

Hong Kong, 30 May – 1 June 2006. A conference devoted to the issue of scalability in information systems, with a theme on information retrieval.
<http://www.infoscale.org/>

Seventh International Conference on Flexible Query Answering Systems

Università degli Studi di Milano Bicocca, Milano, Italy, 7th to 10th June 2006. A query answering conference with a number of themes of interest to IR researchers and practitioners.
<http://fqas2006.disco.unimib.it/>

Desktop Search Tools - Managing the Flight Deck

Bloomsbury Suite, SOAS, London. Wednesday, 14 June 2006. UKeiG Members' Day and AGM, with a focus on desktop search. Please contact Christine Baker, Tel 01969 663749. Email: cabaker@ukeig.org.uk

Joint Conference on Digital Libraries: Opening Information Horizons (JCDL 2006)

Chapel Hill, North Carolina, USA, 11th –15th June 2006. A digital library conference which will be of interest to members working on search in such systems. Also includes a user access workshop.
<http://www.jcdl2006.org/>

Adaptive Hypermedia and Adaptive Web-Based Systems 2006 (AH'06)

National College of Ireland, Dublin, Ireland, 21st – 23rd June, 2006. A general conference on adaptive web based systems including a theme on IR.
<http://www.ah2006.org/>

Visual Categorisation and Image Management Systems (VICMS)

University of Sunderland, St. Peter's Campus, 28 June 2006. A one day multi-disciplinary workshop on categorising images, of particular interest to those interested in image retrieval.
<http://www.cet.sunderland.ac.uk/vcims/index.html>

6th International Workshop on Next Generation Information Technologies and Systems

Kibbutz Shefayim, Israel, 4th to 6th July 2006. A general conference on promoting new research in information systems, including information retrieval.
<http://mis.haifa.ac.il/~ngits>

International Workshop on Intelligent Information Access, 2006 (IIIA-2006)

Marina Congress Center, Helsinki, Finland, 6-8 July 2006. A workshop on uncertainty in information access using such methods as machine learning and probability theory.
<http://cosco.hiit.fi/search/IIIA2006/>

IADIS INTERNATIONAL CONFERENCE e-SOCIETY 2006

Dublin, Ireland, 13th to 16th July 2006. A conference for members who are more interested in the 'soft' side of information seeking and retrieval.
<http://www.iadis.org/es2006/>

4th International Workshop on Adaptive Multimedia Retrieval (AMR 2006)

University of Geneva, Switzerland, 27th –28th July 2006. A specialist workshop on Multimedia IR.
<http://viper.unige.ch/amr2006/>

THE THIRD CONFERENCE ON EMAIL AND ANTI-SPAM (CEAS 2006)

27th –28th July 2006. A Conference of interest to those members who work on adversarial information retrieval. <http://www.ceas.cc/>

29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval (SIGIR 2006)

Seattle, USA, 6-11 August 2006. The world's leading information retrieval conference.
<http://www.sigir2006.org/>

Seventeenth International ACM Conference on Hypertext and Hypermedia (Hypertext 2006)

Odense, Denmark, 23rd –25th August 2006. A conference which will be of interest to members who work in the area of search by browsing.
<http://www.ht06.org/>

ESSIR 2007 - European Summer School in Information Retrieval

27-31 August 2007, Glasgow, Scotland. This five-day school (Monday to Friday) will give participants a grounding in the core subjects of Information Retrieval. www.dcs.gla.ac.uk/essir2007

3rd International Workshop on Text-Based Information Retrieval (TIR-06)

Riva del Garda, Italy, 29 August 2006. A one day workshop on text retrieval using AI methods, in what can only be described as a very beautiful location. <http://www.aisearch.de/tir-06/>

6th International Conference on Knowledge Management.

Graz, Austria, 6th September 2006. With one day workshops on the issue of knowledge organisation and information retrieval and visualisation in IR. <http://i-know.know-center.tugraz.at/>

The 20th British HCI Group conference in co-operation with ACM (HCI 2006: ENGAGE).

Queen Mary, University of London, 11th –15th September 2006. A well regarded HCI conference, which will be of interest to members who have a particular interest in human factors in IR. <http://www.bcs-hci.org.uk/hci2006/>

10th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2006)

Alicante, Spain, 17-22 September 2006. A major European digital library conference, which also hosts the CLEF 2006 cross language IR workshop. <http://www.ecdl2006.org/>

The 17th European Conference on Machine Learning and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (EMCL/PKDD 2006).

Workshop and challenge of Spam filtering. Berlin, Germany, 18-22 September 2006. A TREC like evaluation workshop on Spam filtering held at a dual AI conference. <http://www.ecmlpkdd2006.org/challenge.html>

4th International Conference on Intelligent Information Processing (IIP2006)

Adelaide Australia, 20th –23th September, 2006. A broad based AI conference on information systems, with a theme on information retrieval. <http://www.intsci.ac.cn/iip/iip2006/>

Book Review:

Web Content Delivery, by Xueyan Tang, Jianliang Xu, Samuel T. Chanson

Reviewed by Andrew Neill



What is not immediately clear from the synopsis on Amazon is that this book is a collection of academic papers. These have been written by leading researchers around the world and assembled by the editors as a representation of the state-of-the-art in

web content delivery. As such, the reader should be aware that this is rather dry read, and is definitely not targeted at the 'intelligent layperson'. The Idiot's Guide to Web Content Delivery this is not!

The book's stated aim is to be comprehensive and up-to-date. This is always a dangerous path to take, given the frantic-pace of information technology, and it is difficult to predict which advances will be the hot-topics of the next 6 months. An examination of the papers covered shows that peer-to-peer, wireless networks and utility computing are covered here, all of which are current (or at least recent) threads of discussion in the IT community. Take, for example, Microsoft's recent announcement that its future business model will be to "software as a service" over the internet, the BBC's trials of peer-to-peer downloading of TV and radio repeats, or recent announcements of legalised peer-to-peer networks (such as the upcoming Grokster 3G and Sony partnership with Playlouder MSP). It is clear that this book contains a sample of the research behind these headlines.

The book contains seventeen papers, split into four fairly distinct sections – Web Content Delivery, Dynamic Web Content, Streaming Media Delivery and Ubiquitous Web Access. As an illustration of the type of papers contained in this book, there follows a selection of titles (the full list is available online).

The first section contains "Web Workload Characterisation" and "Content Location in Peer-to-Peer Systems"; the second (and shortest section) contains "Utility Computing for Internet Applications" and "Proxy Caching for Database-backed Web Sites". The third section includes "Streaming Media Caching" and "Policy-based Resource Sharing in Streaming Overlay Networks", and the fourth holds "Distributed Architectures for Web Content Adaptation and Delivery" and "Web Content Delivery Using Thin Client Computing".

As may be apparent from the titles above, there is more here for the serious technician than for the casual browser, and this is more likely to be on the shelves of academics, students, and research and development departments rather than the IT support team. The reader will be following the relative merits of alternative *cost functions* (in "Replica Replacement and Request Routing") or *popularity functions* (in "Peer to Peer Assisted Streaming Proxy").

"this is rather dry read, and is definitely not targeted at the 'intelligent layperson'"

The relevance of the papers, however, should not be underestimated. For example, Netgear recently lost a class-action suit in the US, where a group of customers had sued on the basis that Netgear's wireless routers could not provide the bandwidth claimed in the marketing materials. The paper on "Wireless Web Performance Issues" in the section on Ubiquitous Web Access illustrates exactly the sort of problems that Netgear's marketing department fell foul of. Although bit-rate throughput may be high (and there is great difficulty in measuring exactly how high, or defining the measurement conditions), actual bandwidth is reduced because many of these bits transferred are protocol messages, not meaningful data. This paper quantifies the various bottlenecks in wireless content delivery, and urges further research - something that Netgear should certainly heed!

The book is professionally produced, clearly printed and well laid out. The black and white illustrations are adequate, but plainly scientific. Although bound in hard-cover, it is reasonably sized for easy transportation, so could be a companion on a long journey to a conference. There are a couple of minor misprints, but no meaningful mistakes, and nothing that reduces the impression of a carefully assembled package.

It is fair to say that the goal of summarising the state-of-the-art has been achieved, and there are plenty of references for the interested reader to follow up on. Again, be warned that this is not for the casual reader. Only the specialists in the field - those involved in commercial research labs or universities - will be able to make the most of the information. Those who work in IT support, unless highly motivated, will find little (though not nothing) of practical use. Better to wait for the boffins of this world to capture the genius lurking within these pages and wrap it into the latest CISCO router or Apache web server. It should certainly grace the computer-science shelf of any self-respecting university library.

Andrew Neill is a Business Analyst within the Document and Knowledge Management team in international city-based law firm Norton Rose. Prior to this, he worked in the technology and integration division of Deloitte & Touche Business Consulting, working on a wide variety of technology projects for blue-chip companies. Andrew also recently completed an MSc in Computing for Industry at Imperial College. He specialises in knowledge and content management systems, enterprise and service-oriented architecture, and business process improvement. Outside of work, his interests include languages and travelling, and he and his fiancée like to spend as much time as possible enjoying friends' hospitality, particularly in Spain, Italy and Greece. He can be contacted via:
andrew.neill@nortonrose.com

Feature Article:

New Players in the Information Society

By Theo Huibers



In a recent study the Dutch Social and Cultural Planning Office speaks of “the multiple choice society”. Citizens now have more choice than ever. And this freedom of choice means they require information to make the right choice. If you need a mortgage, you can choose from countless products

and suppliers, each with their own specific terms and conditions. And if you’re looking for a home, a holiday or a restaurant, you can take your pick from a huge amount of options.

Faced with all that choice we’ve developed an almost insatiable appetite for information. Choosing is becoming increasingly complex and so we also need more and more (background) information and guidance. In our services economy these developments have sparked an enormous increase in the number of intermediaries active in such fields as insurance, home finance and travel. They serve up the vast array of product offerings in bite-sized chunks so that you can digest all the facts and make your choice. In other words, they give you the information you need to make the best buying decision. Publishers are also well-placed to make hay in this intermediary world. Their business is not to select holidays, mortgages or insurance, but information. On the basis of their selection, they provide information to their customers who are thus able to make better decisions.

Such intermediaries have set up a business model around the advisory process. Tasks are performed by individuals with (powerful) support from information systems. The intermediaries exist by virtue of their ability to collect, select, process and enhance information in a specific way for a (specific)

target group; information that everyone can in principle access, either free or for a fee. Everyone, of course, can shop around for the best or cheapest mortgage. But the advantage of intermediaries is that they have made this their field of expertise. They keep up with the latest offers and innovations on a daily basis and make this multitude of information digestible for their customers.

Thanks to new technologies information can now be automatically collected, selected and, to a certain extent, enhanced. These systems are also known as information retrieval systems. Familiar examples are the popular search engines Google, Lycos and Yahoo!

This new way of collecting information is taking a heavy toll on the traditional intermediaries. Now that travel information, for instance, can be easily obtained via the internet, the added value of a travel agent as a source of information has become marginal. Publishers too are said to be losing their unique position in the information society for the simple reason that relevant information is now readily available free of charge.

This article homes in on all these changes. First, we will analyse the unavoidable rise of new (automated) information intermediaries. Next we will discuss the impact of the existing information-intensive value chains. The changing environment also compels us to alter our view on the existing income model. Finally we will look at some new players who may soon be playing a central role in the information society.

The Relentless Rise of New Information Intermediaries

Information technology, and particularly the internet, has given us easy access to an overwhelming amount of data and information. The floodtide of information is unstoppable. It has been said that on an average weekday the New York Times contains more information than a person living in Shakespeare’s day would have gathered in his entire lifetime. Via the internet we can access not one but thousands of digital newspapers.

In order to keep the rapid proliferation of digital information under control, we need

automated systems that are able to search for that information on the basis of relevance. In the past decades many practical information retrieval systems (IR systems) have been developed. Such systems can be used as an information filter (e.g. the popular spam filters) but also as a search engine (e.g. Google, Ilse or Yahoo!). The available systems are becoming more and more sophisticated and are increasingly capable of satisfying the user's expectations.

The information society and commerce have obviously embraced information retrieval in a big way. Google, for instance, provides access to more than three billion web pages and almost the entire internet population makes use of search systems. Another significant fact is that 70% of all eCommerce-transactions are initiated by an online search. Prospects and customers are less likely to find your organisation via www.yourcompanyname.com than by using a search engine.

Changing Value Chains

Information intermediaries are now repositioning en masse or are being replaced by other players. In short, the entire value chain of the information process is in a state of flux.

Information chains change and consequently so do the accompanying value chains. This has huge consequences for the traditional intermediaries who must change to survive. Those who remain stuck in the old pattern will inevitably go under. The really successful players in the new chain are usually not the familiar old players in a new outfit, but relatively young companies like Google and Overture. It's not the selling of information but the search for information that has become big business.

The challenge for both old and new information intermediaries lies in building effective filters and search engines to connect supply and demand. But how can you earn just as much doing that as you could in the old paper-based world where the selling price basically consisted of the cost price plus a fat margin. In the digital environment, "relevance" will be the key price-determining factor. In other words, the value of information is largely related to the relevance of information.

So the bottom-line question is: how do you calculate the costs of relevance? The degree of relevance must somehow be expressed in financial terms. This depends on a combination of multiple factors, including up-to-dateness and reliability but also the receiver's needs and existing knowledge. Setting the right price for a service is one of the most challenging issues for the coming years.

The Future of the New Players

This article showed how existing intermediaries who focus on collecting, selecting, processing and enhancing information are being or will be confronted with severe competition from new entrants. The question remains who are the new players and how robust is their future.

At first sight it would seem that owners of internet search systems can look forward to a lucrative life as the intermediaries of the future. You could say that companies are almost compelled to do business with them as their search systems determine whether the consumer will find the company's website. In this way they can create a virtual monopoly in the intermediary market. But the life of an internet search operator is also full of uncertainties.

For it is still impossible to predict which internet search system will prevail in the near future. User convenience, speed and the size of the accessible collection of documents will probably be the deciding factors. But, as Google has shown, the ability to rank hits in order of relevance is also vital. In the space of one year, Google grew from a relatively unknown internet search system into the absolute world leader thanks to their brand-new method of detecting documents according to relevance. The innovative part is that Google determines relevance partly on the basis of a site's popularity, which can be measured by the number of referral links and the referral ratings. The big question, however, remains: will a new IR system be able to poach Google's entire customer base? A more advanced IR system might be able to pull off such a coup, and so might a system that is integrated into a control system or a standard company application.

The second group of new players are those who offer their services on the basis of existing

search machines. Think, for instance, of internet travel providers who are slowly but surely driving travel agencies out of business. The same is happening in the advertising, car, housing, investing, mortgage, airline and insurance markets, to mention but a few examples. The future of these new intermediaries is also far from certain. The customer's loyalty to an internet-based intermediary is minimal. What's more, online customers are always free to shop around for a better offer before deciding to buy. Time will tell whether the customer is also prepared to pay for the quality and relevance of the information supplied on the services offered. If so, we may see the birth of a new generation of agents who collect information which only paying consumers can access. In such cases the intermediary will collect a fee for his services, which brings us back to the traditional business model of publishers.

Finally, new players may also emerge from the old intermediaries of today. Publishers, for instance, have for some time been seeking to secure a position in the new digital information landscape. They are trying to win the loyalty of customers by offering better access, good-quality information and a select range of products. But can existing intermediaries make this jump into that new role? That is by no means certain. For it means redesigning their entire business model and everything related to it.

In Conclusion

The position of traditional intermediaries is under fire. The rise of advanced IR technology makes it possible to introduce new business models. Collection, selection, processing and enhancement of information can be taken over by automated systems. This leads to new business opportunities for new players in the information society. It will drastically change the original business models and value chains of existing information intermediaries - and possibly all intermediaries - in our services economy.

Prof. dr. Theo Huibers MMC is Professor in Information Retrieval at the University of Twente and also director at KPMG Business Advisory Services (www.theohuibers.nl). He can be contacted via info@theohuibers.nl

Product Update:

Trexy.com: New Search Trails Technology Inspired by a Vision Way Back in 1945

By Megan Hamilton



Trexy.com is a new search technology, inspiration for which dates way back to a visionary paper written in 1945.

Trexy.com is inspired by scientist and visionary, Dr. Vannevar Bush, who

in the 1940's shared his vision of creating a collective memory by recording peoples' trails through information. Bush shared this vision in his paper "As We May Think."

<http://www.theatlantic.com/doc/194507/bush>

Bush dreamt of a device called a "Memex" – an enhanced supplement to personal and community memory created by trailblazers: *"those who find delight in the task of establishing useful trails through the enormous mass of the common record."*

Bush's article influenced Ted Nelson (father of hypertext) and Tim Berners-Lee (creator of WWW), who have each realised a part of Bush's Memex vision.

Trexy's CEO and Inventor, Nigel Hamilton said: "Our brain records by burning memory engrams or trails between associated neurons. Each time a memory trail is traversed the synaptic gaps fire between neurons and the engram is reinforced. The more times this trail of neurons fires the stronger the memory becomes. If it fires less frequently the memory fades".

"Imagine if your trails didn't fade?" he said.

London based Trexy.com aims to fulfil the trailblazing aspects of Bush's vision by enabling users to create search trails. Search trails are the pathways users create while searching on their favourite engines like Google, MSN and Ask Jeeves.

"Every day we all make personal discoveries and traverse trails from what we're looking for, to what we find - but sometimes we forget. What did I search on again? What was I looking for? How did I find that?"

"We're all in an endless loop of searching, finding, sometimes forgetting and the procedure repeats. Wouldn't it be good if you could harness all this effort? Wouldn't it be good if you could remember your personal search trails and also follow the search trails of others?" he said.

Trexy.com remembers the search terms and the web pages visited on over 3000 engines. Trexy also enables users to follow the anonymous search trails of other searchers. "Instead of searching for the same thing twice, Trexy helps you to pinpoint information you've already found. So you'll never have to ask yourself, "Where did I find that again?" Nigel Hamilton said.

Search trails can be created at Trexy.com or by downloading a free toolbar called the Trexy TrailBar (<http://trexy.com/trailbar.html>). Trexy is free to use.

Megan Hamilton is Director and CMO of Trexy. In addition to her legal qualification as a barrister, Megan holds a masters degree in international management and a bachelors degree in business communication. Prior to establishing Turbo10 and Trexy, Megan has worked as a marketing lecturer. She can be contacted by email via: media@trexy.com

Interested in Search Engines? Based near London?

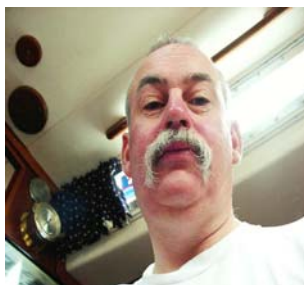
The Flag and Bell is a Tech Pub Crawl held on the second Tuesday of each month. It is a free, networking event for anyone interested in search engines, the Internet or web technology. A free drink at each pub will be sponsored by Trexy.com.

For further details, see <http://flag-and-bell.com/>

Feature Article:

A Tribute to Karen Sparck Jones

By John Tait and Mounia Lalmas



Karen Spärck Jones reached the age of 70 on the 26 August 2005, and an event was held at Downing College, Cambridge, on 5 April, which launched a [book](#) as

an early celebration of this occasion. The day consisted of a series of talks, the reading of a poem, the showing of a video, and, at the end of the day, a celebratory dinner. 45 people attended the day, with four travelling from the United States especially for the event.

The day was originally conceived as a series of talks by authors of chapters in the book, but in the end it turned into something quite different. The only people who followed fairly closely to the content of (one of) their book chapters was Donna Harman, and to a lesser extent Stephen Pulman and Ted Briscoe. Several people who had not contributed to the book for one reason or another gave talks, and Steve Robertson who contributed a chapter spoke about topics covered in a chapter written by somebody else. Many of the speakers related various fondly (and in one or two cases not so fondly) remembered anecdotes about Karen as well as the usual academic stories. Most of the talks also included some profound reflections on the influence for contemporary research in language and information of the enormous and wide ranging body of work undertaken by Karen over the last forty years or so.

The day began with a talk on noun compounds (e.g. "airport long term parking courtesy bus") by Ted Briscoe, from the University of Cambridge, who presented a number of issues regarding the processing of English noun compounds, with a view to integrate statistical approaches, as these can (at least partially) replace domain knowledge that is very difficult and time-consuming to build.

The second talk was given by Henry Thompson, from the University of Edinburgh, who did a remarkable job, stepping in when Gareth Jones had to withdraw at the last minute, because he was unable to get from Japan to the meeting in time. Henry talked about some recent developments in the semantic web, which he related to Karen's early work. His talk ended by suggesting using statistical methods for semantic web so that to better distinguish between the enormous amount of useful and non-useful information on the web.

Stephen Pulman, from the University of Oxford, then talked about models of word sense, their strengths and weaknesses. Steve discussed established [theories of meaning](#) in philosophy, and related them to the issues of word senses in both computational linguistics and information retrieval research.

The afternoon began with Mark Maybury, from MITRE Corporation, who also could not attend because (in his case) of last minute problems with a conference he is organising, joining us by telephone from the US to read his poem "Ode to a Spärck". This was followed by a video Mark has produced of snippets of good wishes to and reminiscences of many leading Computational Linguists and Information Retrieval people from around the world.

Next was Donna Harman, from NIST, who talked about the importance of focussed evaluation so that to be able to understand how systems (e.g. retrieval, summarisation, question and answering, etc.) perform and to compare different ones. Her presentation concentrated on [TREC](#) (Text REtrieval Conference), and [DUC](#) (Document Understanding Conference), both of which were very influenced by Karen.

Barbara Grosz, from Harvard University, presented her "[Beyond Mice and Menus](#)" paradigm to advance the collaborative aspect of user interface, so that to arrive at truly interactive (useful and helpful) interfaces.

Stephen Robertson, from Microsoft Cambridge gave a talk on [inverse document frequency](#) (IDF), a topic covered in a very different way in a book chapter by Donna Harman, where he explained the relation and the non-relation

between IDF and Shannon theory of information.

The last formal talk of the day was by Candy Sidner, from MERL, who gave an extremely entertaining talk about [experiments with a robotic penguin](#), in an attempt to build an automated participant in human computer interaction with gestures. The day was wound up by Andy Hopper, Head of the Computer Laboratory, Cambridge, who announced how interesting he had found the day.

In the evening there was a formal dinner in the University Centre in a lovely room overlooking the river. Yorick Wilks, from the University of Sheffield, spoke after dinner, looking back at the many things that have happened since he first worked with Karen in the Cambridge Language Research Unit in the 1960's. Karen gave a charming reply, thanking all those involved for the day and the book. The day closed with a toast from Jane Robinson, formerly of SRI, a lady who is old enough to be Karen's mother!

The very strong attendance, despite the rather short notice at which the meeting was arranged, the efforts many people made to attend the meeting – several from Edinburgh for example, as well as those from the US, and the Netherlands: all evidence of the tremendous respect and affection with which Karen is held.

One of the most pleasing things about the day was the number of senior (and not so senior) academic women who attended the day. It is easy to forget that it was not so many years before Karen graduated that women were first allowed to take their degrees at Cambridge. As well as Karen (who some would argue had to wait much too long for her chair) there were five female full professors in the audience. Clearly, things have moved on. However, there were eight male professors present, so still some way to go there.

Anyway, the main thing was that Karen clearly thoroughly enjoyed the day from every point of view. She made many incisive comments in the course of the day. Let us hope we have many more years of Karen's wisdom to look forward to!

Some photographs and other memoirs of the event may be found at <http://osiris.sunderland.ac.uk/~cs0jta/KarenEventMemoire.htm> as well as a form which allows a few remaining discounted copies of the book to be ordered (£35 including postage and packing, as opposed to the Springer Price of £62 plus p&p).

John Tait obtained his Ph.D. in 1985 for a thesis entitled "Automatic Summarisation of English Text" which was supervised by Karen Sparck Jones, with whom he also worked on the use on Natural language Processing for query expansion. He is now Professor of Intelligent Information Systems at the University of Sunderland where he continues to work on summarising, the use of NLP for IR, and also now semantic content based image retrieval. He can be contacted via John.Tait@sunderland.ac.uk.

Get Involved!

Informer welcomes contributions on any aspect of information retrieval. We are particularly interested in feature articles and opinion pieces, but are also pleased to receive news articles, book reviews, jobs ads, etc.

Right now we are running a series of Product Reviews, so if you are interested in reviewing any of the following:

- [Copernic](#)
- [Ask Jeeves Desktop Search](#)
- [Blinkx](#)
- [MSN Search Toolbar](#)

Then please get in touch with us via irsg@bcs.org.uk. All of the above are freely available as software downloads.

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](#).