Informer

Number 15, Summer 2005
ISSN 0950-4974

BCS INFORMATION RETRIEVAL

Contents

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.

Those of us who've been IRSG members for a while will no doubt be aware of the group's history, with its traditional focus on academic research. After all, IR is a largely academic discipline, isn't it? Well, it may have been in the past, but things are somewhat different now. In fact, search is all around us, and it's big business. As many of you will know, Microsoft has been investing heavily in search for the past couple of years, growing both its R&D and product teams, and promoting its own search services and technology. Most notable of these is its own web search engine: MSN. It'll be interesting to see how this fares against the three main web search incumbents: Google, Yahoo and Ask/Teoma. Browser wars all over again?

But of course it's not just about searching the web. In fact, there is something of a battle going on much closer to home, with Google trying to leverage its web search technology to occupy a space that MS should really have sewn up years ago: desktop search. You can read more about Google's Desktop Search offering on p2, in the first of a new series of product reviews. (And incidentally, if you are interested in evaluating any of Google's desktop search competitors, such as Copernic, Ask Jeeves Desktop Search, Yahoo Desktop Search, Blinkx or the MSN Search Toolbar, then see page 12.)

However, automated search solutions are only part of the answer: there are still some things that are best done using manual judgement and expertise; web site indexing being one of them. Heather Hedden tells us more about the role of human indexers on p5.

One other positive aspect of the current investment in search is the career opportunities it presents – to the extent that for this edition of Informer we had hoped to run (for the first time) a jobs section. But the problem isn't *finding* the content to fill this

section – it's *filtering* it down into a form that makes sense to our readership and complements the rest of the material (a classic IR problem you might say!) There are simply too many IR opportunities out there to just list the ones that we're aware of, so if you have any thoughts on how we might prioritise them, or what kind of jobs section you'd like to see, please let us know.

Likewise, our Events Section seems to get longer with every issue! Evidently, the Summer season is always a busy time in any conference calendar, but with the listings now spreading over two pages, perhaps it's time to re-think how we use our various publishing channels, and perhaps whether the IRSG website might not be the better vehicle for publishing such "reference" material. Again, let us know your preferences – just drop us a line at irsg@bcs.org.uk.

Finally, a word about book reviews. If you were one of those who were disappointed not to receive a copy of the Semantic Web book we offered recently, not to worry. We have made an arrangement with Springer to review their books on an ongoing basis, and any announcements on upcoming titles will be made here first (see p11 for further details). So be sure to get your copy of Informer as soon as it comes out to have first pick!

Best regards,
Tony Rose
Editor, Informer
Email: irsg@bcs.org.uk

Product Review: Google Desktop Search

By John Dumas

It's got to be in there somewhere...

We all know the problem. You are convinced, absolutely, positively, that an email or file or image is on your computer somewhere. But try as you may, it becomes impossible to find. You continue to look and, like lost house keys, you get more and more frustrated and more and more convinced it should be in the places you are looking. You try Windows XP search and end up cursing Microsoft for not using a mature and experienced retriever. Your children take offence as you shout insults at the puppy dog. You find yourself drinking a whole pot of tea while you wait for Outlook advanced search to, apparently, index your whole email system each time you try the first query. You begin to have visions of confronting Bill Gates… "Cripes man! It's just plain text!"

Getting started

During installation, the initial indexing is painless enough. There is an option to continue working as normal, allowing the indexing to progress in the background, and even to pick-up later if you turn off your computer. However I chose to run this initial phase like a virus scan, giving the computer full resources, letting it cook away on its own. The time this takes will depend on the amount of material to be indexed but even for my relatively deep and populated file space, with over 70,000 items to index, it felt efficient and far faster than a full virus scan.

One of the key reasons I wanted to try a desktop search tool was the poor experience I have had with the built-in search for Outlook, email in particular and especially messages I have archived into separate Personal Folders Files (.pst). So I took the time to read some

of the installation help and realised I needed to have the relevant email clients and archives opening and running during the indexing phase. This meant I had to have Thunderbird, for my personal email, running and all my current and archived Outlook PST files open and running in Outlook, for my work email.

Although after installation you have control over what to keep in the index, via preferences, I had concerns it would cache and index my online banking activity. Adding web domains you want to omit and always ignore is very straightforward however I would question the level of awareness and understanding of this for more novice users.

How good is it?

Because it uses a dedicated lightweight index, the speed at which you get back results is far superior to any of the native Microsoft search options.

Relevance of results is very good and lives up to expectations set by Google Web Search. In fact there is even a feeling that results are superior to Google Web Search. You are searching your own restricted domain space and you know the right kinds of keywords to use and know the right kind of scent to pursue. So it at least feels as if you reach your goal faster.

Images suffer from their typical lack of meta data however if you have provided a relevant keyword in either the file or folder name, then results appear to be acceptable. Thumbnails of images help you visually scan results for relevance. Frustratingly there is not a filter provided to display results by media type, though you can filter results by 'email' or 'file'. Since they provide a nice thumbnail feedback for images, it would be nice to filter by images only. It also appears it does not index or search within compressed/zipped files, which is problematic when I want to find a document in my archived, i.e. Zipped, folders.

There is no noticeable overhead with Desktop always running in the background on the three different platforms I use (laptop, home and work desktops). Though all of these have fast processors and 512MB of RAM I feel this is a reasonable indicator since this specification is rapidly becoming the baseline. However I do love lightweight and efficient software so it would be great to see if this holds when running it on a thinner client.

Real time indexing is uncanny, what you see is what and when it's indexed. That shouldn't be surprising but with no overhead on resources, you forget it is even happening. A quantum leap from Windows and Outlook search options.

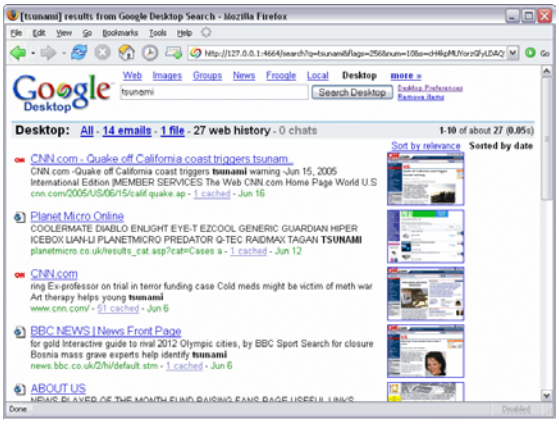
Using it

A clear usability advantage is that the interface, results format and options for searching are almost identical to Google Web Search. So using it is a snap since most of us have already developed a level of Google expertise.

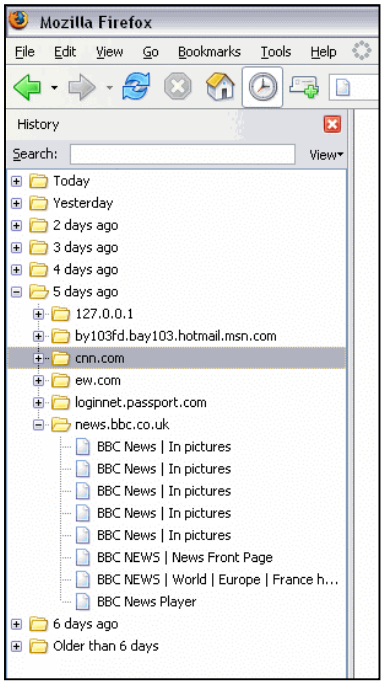
Results are displayed as 'Sorted by date' by default, but this seems to fit with my typical way of organising my files and emails so I rarely feel the need to select 'Sort by relevance'.

Once the desired result is found, the 'Open folder' link allows a nice transition to the Windows Explorer/My Documents view so the document can be used as normal via the OS environment. At which point I usually think, "what was I thinking when I put it there?"

I have yet to break my habit pattern of using the browser history feature for finding my way back to Web pages I recently visited. Because there are times when I use both Firefox and IE, this can be awkward if you forget which browser you were using. The 'web history' filtering option appears a promising alternative. It provides a nice thumbnail image of the web page and sorts by date, most recent first. But it will be interesting if



this is as useful as the expandable folder view in Firefox history where it can display sites by date and then group them by domain. There are additional third party plug-ins for Google Desktop, which I have yet to explore, and these may well add additional 'sort by' options and 'expand/collapse' views.



The online help is a bit sparse and possibly cryptic in places for novice users. It's also hard to find since you have to go through 'Preferences' first. However a quick scan makes me realise there are additional search operators you can use, especially for searching emails, which overcome some of the options missing in the interface I previously mentioned. For example, it surprisingly fails to explain that you can search for media types by adding wildcards like "*.gif" to the keywords or phrase. This never occurred to me since you feel as if you're in a Web Search mode where file specific search operators are typically not a practical technique. It would be nice if, once you choose to filter by "files", some contextual help appeared to show how wildcards could further refine the search by document or media type.

Who's watching?

There are two areas of privacy concerns I had with a tool as powerful and as easy to access as this. Firstly, is Google data mining my personal activities, even email and personal files? Secondly, does this tool make it frighteningly easy for other users of my computer to access personal communications, information and activities?

The press flurry around Google's intention to data mine Gmail user email triggered my immediate suspicion when installing the tool. Though it claims it will only send 'non-personal data' and 'crash reports' I found myself paranoid from the Gmail strategy and was on the lookout from the very start to deselect any setting which would send back reports. There is just something unnerving about the idea that a report related to all my emails and files would leave my machine for someone else's use. If Google could simply provide some further assurances and explanations, I may likely soften up on this issue in the interest of helping them to develop a better product.

Regarding the second area of concern, I suppose I shouldn't really worry. I am boringly prudish in my online activities and just don't seem to have the time or energy to engage in contentious electronic gossip or scheming. But much of my personal and professional life is spent communicating through email and using the web. So I do find it slightly worrying that, because Google Desktop is so powerful and effective, with only a few clicks someone can find out anything they might be interested in about "my virtual space".

Finally, the default setting to integrate Desktop results with Google Web Search results straddles both of these areas of concern. With local results mixed with web results, it immediately feels as if somehow my information might be sent to Google, despite my reasonable knowledge of how web technologies work. And of course another user of my computer would easily stumble across my files and emails while innocently performing a Google Web Search. Therefore I immediately turned off this feature as well.

Satisfied "customer"

A search approach is not my first port of call to find a file or email because I always try to take the time to store them in a hierarchical folder structure. So I find my need to use search has always been infrequent. However, like those lost house keys, when you need a fast, easy and powerful way to find something, Google Desktop is the biz! And the price is right.

John Dumas is Director of Operations at optimum.web ltd., where he has led user

centred design projects for a diverse range of clients, such as Microsoft, Land Rover, Diageo, Amway, and the DTI's Small Business Service. He has lectured in the UK on HCI, internet applications, e-commerce and digital media. John has over ten years experience in usability and web development in the public and private sector, both in the UK and USA. With a love of document structure and web standards, he enjoys specialising in web accessibility and, sadly, takes pleasure in reading W3C Recommendations. He can be contacted via: ***john@optimum-web.co.uk***

Feature Article:

"Enhanced Site Searching with an A-Z Index"

By Heather Hedden

As web sites have grown, so has the challenge to help users find information within a site. There are two ways that a user may look for information within a site: navigation and searching. Navigation is the exploration of a site to find out what information is available. Searching is the function of trying to find information on specific topic. In recent years the specialty of information architecture has emerged, which has significantly improved site navigation. Improvements in search capabilities, however, are still needed for many sites. Because the web is a relatively new medium, new techniques and technologies, such as search engines, tend to be applied. But traditional methods of searching, with slight modifications, such as the A-Z index, can also work very well in serving the search needs on web sites.

Drawbacks of Search Engines

The most common method to enable user searching of site is by adding a site search engine. Search engines do have significant drawbacks, though.

Site search engines may not retrieve enough or any pages. Search engines for the entire web usually produce satisfactory results in the quantity of pages, as users generally want "some information about" a subject, and this can typically be found on some of the numerous pages retrieved. If many good pages are missed by the search engine, the user usually does not notice or care. Within a web site, however, the number of pages is relatively small, so a simple search engine search might not yield enough or any results, even if there are good pages on the subject. This is most likely to occur because the search subject that the user enters is worded differently from the references to that topic within page text.

Site search engines may retrieve too many irrelevant pages. Web search engines usually produce satisfactory results in the quality of articles, since the major search engine companies have developed complicated criteria and algorithms for the retrieval and ranking of pages. The search engines to be used within a site are not so sophisticated. They often retrieve pages that include a mere passing mention of the search term, but do not really focus on the subject at all. In the end, the quality of the search engine results reflects the sophistication of the search string entered by the user, which cannot be controlled. In the A-Z index, on the other hand, the quality of the results reflects the sophistication of the indexer.

Site A-Z Indexes

A-Z indexes are created not by machines, but by humans who take care to add index entries only to pages on which good information about the topic appears. In this way, the indexing of topic words mentioned in passing or out of context is avoided, boosting the overall relevance and quality of the index itself.

An A-Z Index offers an alphabetical list of "entry point" topics through which the user may browse and select. In an index at the back of a book or manual, the entries are followed by page numbers. On a web site, the entry points are hyperlinked to the appropriate pages, and often to named anchors within web

pages for an even greater level of detail in indexing.

As with book indexes, a site index may contain multiple entries, each worded differently, that point to the same page, or page and anchor. This approach is used to cover all the different ways a user may think a topic is named. Indexers call this feature "double posting." It covers synonyms, such as "cars" and "automobiles," and the different word order of a phrase, such as "automobile engines" and "engines, automobile." The browsable nature of the index solves the problems that might arise from incorrect or variant spellings, and singular vs. plural usages that the site user might choose. In addition, there is often a second level of terms, called "sub-entries," that are listed and indented under some of the main entries.

A web A-Z index is typically a single, long HTML page, although it could be broken into separate pages for each letter of the alphabet if it were long. At the top of the page, a horizontal list of the letters of the alphabet usually appears. The user makes a selection from this list, and jumps to the appropriate section of the alphabetical index.

A list of some examples of A-Z indexes can be found on the Web Index Examples page of the Web Indexing Special Interest Group: http://www.web-indexing.org/web-index-examples.htm

Site Indexes versus Site Maps

Site maps are not an alternative to site indexes, since they act as the table of contents and serve a somewhat different purpose. But some web site owners and designers are unaware of the difference, and might even mislabel a site map as a "site index."

A site map tends to reflect the hierarchical structure of the web pages of a site with categorized web page titles. A purpose of a site map is to have a list of a site's web pages that can be quickly scanned in one screen view with minimal scrolling, without having to go through each menu and submenu one by one. A site map might not include all the pages of a web site. If the site is large, only the top few levels of the hierarchy would be displayed. The entries in the site map tend to be the page titles, but they could be modified slightly. Each page in the site map is listed in only one place. Like the navigation menu, the site map is to aid navigation rather than searching, and thus should not be seen as an alternative to an index.

There are software tools to aid in the creation of site maps, by extracting web page titles along with their hierarchical links within a site. The problem is that some of these tools also offer the feature of alphabetically sorting web page titles to create an "index," a feature that probably should not have been created. An alphabetically sorted table of contents in a book does not create the book's index, so this method should not be used for web site indexes. An alphabetized list of topics or names is useful; but a list of alphabetized page titles is not.

Site Indexes versus Taxonomies

Taxonomies are hierarchical classifications of terms, concepts, or topics, in a tree-like structure. On a web site taxonomy, the user typically clicks from one level to the next most specific level. An example of a very broad taxonomy is on the Amazon.com site http://www.amazon.com in the left-hand bar of categories under Browse.

A taxonomy's function and purpose falls somewhere between that of a navigation menu and that of an index. As with an index, the terms or labels of a taxonomy tend to be carefully selected, and a narrower concept can be placed in more than one place in the taxonomy. The purpose of a taxonomy is more that of guided search, than for site navigation.

Certain types of sites or parts of sites are better served by taxonomies, and others are better served by indexes. If most of the content is dealing with a narrow subject area, such as a web site devoted to information on a product line, heart disease, or historic films, a taxonomy might work better than an index. An index, on the other hand, serves best a site with varied types of content. There has been a lot of interest recently in taxonomies for aiding the organization and retrieval of information in large web sites and intranets. As a result,

taxonomies might end up being implemented where an index would actually serve better.

Site Suitability for an Index

The most suitable sites for A-Z indexes are those with repeat visitors (intranets, government sites, organizations, periodicals, and companies offering recurring services), of a medium size (perhaps 30 to 600 pages), with not too many pages changing too frequently, and with a rich and varied content.

Who does Indexing?

Creating an index is more complicated than creating a hierarchy of categories. To become competent at indexing requires appropriate training. Information architects with backgrounds in library science and a good sense of labelling, however, could probably pick up indexing from reading a good book on the subject. In addition, a tool for automatically embedding the index is recommended.

If you don't want to invest the time and energy in learning indexing yourself, it's probably best to contract a freelance indexer. Most of the professional associations of indexers, listed below, maintain searchable databases of freelance indexers. Limit your search to HTML or web indexers. The nice thing about web site indexes is that samples of an indexer's work are usually accessible online, so you can easily evaluate a potential indexer's work.

- Society of Indexers: http://www.indexers.org.uk
- American Society of Indexers: http://www.asindexing.org
- Indexing and Abstracting Society of Canada: http://www.indexingsociety.ca
- Australian and New Zealand Society of Indexers: http://www.aussi.org

There is also a database exclusively of web site indexers on the Web Indexing Special Interest Group of the American Society of Indexers: http://www.web-indexing.org/contract-indexer-search.php

Conclusions

Search engines are often indispensable on a very large and changing site. Yet a site certainly can have both an A-Z index and a search engine. This would make sense if certain pages of the web site are skipped in the index or not indexed in detail because they are constantly changing.

As web sites, and especially intranets, tend to be large and complex, different kinds of indexes on different parts of the same site may be the best solution for the site. An A-Z index or a directory could be applied to just the top few levels of pages of the site for overall site navigation, while a database could be applied to a section of numerous frequently updated articles. A thorough A-Z index could be applied to a book-like section, such as a policy manual. Finally, a taxonomy could be applied to collection of similar pages dealing with various entities, such as a product directory.

Just because A-Z indexes follow an older style, that of book indexes, does not mean that they are too outdated to serve web sites well.

Heather Hedden, principal of Hedden Information Management (www.hedden-information.com), is an information architect and indexer with a specialty in web A-Z indexes. Previously she worked on thesauri and taxonomies as senior vocabulary editor at the Gale Group. She is vice president of the New England Chapter of the American Society of Indexers and coordinator of the Web Indexing Special Interest Group. She can be contacted at heather@hedden.net.

My PhD:

"Adaptive User Interfaces: Enhancing Interaction by Modelling Relationship Factors"

By Fiona Walsh

Background

Despite what a growing, widespread popularity among users might otherwise suggest, mobile devices are still not without their problems. In particular, the interaction between a user and their device is, typically, the most prevalent concern documented. [http://portal.acm.org/citation.cfm?id=371920.372181]. This issue, commonly attributed to restricted hardware specifications i.e. small screens, limited input/output modalities etc., has been subject to much research and investigation over the last few years.

Predominately, researchers in the past have considered the user and their device as the most important elements involved in the interaction and respectfully so. However, only minor consideration has been given to the information which is also a significant factor in an actual exchange (or interaction).

In this research we aim to give that overdue consideration in a bid to improve the multifaceted entity which is interaction.

Adaptive User Interfaces

Adaptive User Interfaces (AUI) are currently being held up as one of the possible bids to address the HCI challenges associated with user to mobile device interaction. By default, their purpose is to serve and support a diversity of users with their ever-changing information needs and tasks in mobile environments.

Past work in the area of AUIs has benefited greatly from a number of disciplines (IR, User Modelling), who have assumed a number of approaches to solving these challenges. One particular trend, has been to keep the GUI static but to adapt its content using information retrieval and filtering techniques, which some researchers believe to be essential for coping with users' ever-changing information needs in mobile environments [http://www.dcs.ex.ac.uk/~pjbrown/papers/sigir2002.pdf].

Another type of AUI, popular with researchers, is one which dynamically changes specific elements of itself i.e. presentation mode, in order to perform more efficiently in its current environment. This type of context-aware interface is sensitive to a collection of variables such as user preferences and contextual factors.

User Modelling

User Modelling has long been used in conjunction with AUIs [http://www.ilrt.bris.ac.uk/publications/researchreport/rr1085/report_html?ilrtyear=2004]. Analogically, user modelling could be viewed as the inner workings of a clock while the AUI is its face.

Typically, models of the user and/or task are generated implicitly i.e. observing user behaviour, or explicitly i.e. user input; and are used to tailor or personalise aspects of an application's GUI for particular users.

Research Aim

This research plans to address the problem of poor satisfaction levels with regards to interacting with mobile devices by, initially, modeling the relationship which exists between a user, their mobile device and their information; and then using it to tailor the user's interaction with their device.

Conclusion

In spite of all the research effort in AUIs and related disciplines, we are still not where we want or need to be in terms of moderately flawless adaptation; simplified, natural interaction; and most importantly, user satisfaction. We therefore need to build on what has been established, implemented and tested before us; and bring research in AUIs to the next level.

The aim of this doctoral research is to improve interaction between users and their devices and indirectly their information by developing

an AUI which is sensitive to relationship factors as well as user preferences and contextual elements. These relationship factors, we deem just as important to achieve more effective and natural, user to device interaction.

Fiona Walsh is a Software Development graduate (2003) from the Galway-Mayo Institute of Technology, Co. Mayo, Ireland. She is currently pursing a PhD in Computer Science at the Robert Gordon University, Aberdeen under the supervision of Dr. Ayse Göker. Her research interests are HCI, IR and User Modelling. She can be contacted via: fw@comp.rgu.ac.uk.

Workshop Report:

Evaluating User Studies in Information Access

By Alex Bailey

This June a workshop was held as part of the CoLIS5 conference in Glasgow that brought together researchers interested in user studies in Information Access. The aim of the workshop from the outset was to promote discussion and to hold a genuinely interactive workshop that will benefit the community.

User studies are becoming more and necessary to be able to assess the true impact of a technology, especially in a field such as Information Access where the systems developed are ultimately for the benefit of the people using them. It is only with real user studies that the impact of state of the art research can be truly assessed and the merit of such research validated. However, the complexity in designing, running, and analysing a user study is substantially more time consuming and challenging than a simple comparison of empirical measures such as precision and recall. As a result many researchers shy away from the user studies. As to how a user study should be performed in the context of Information Access remains a challenge, and those researchers wishing to perform such a study are faced with many issues to ensure that the research is carried out in an appropriate and unbiased manner.

The broad aims of the workshop were to bring together both experienced researchers and those new to the field to assess current user-based studies in Information Access and to provide a forum of discussion for proposed user studies. The ultimate aim is to draw up a list of guidelines, methodologies and resources for future user studies.

The workshop was planned from the start to promote discussion. A call was made for both experience papers from past user studies, and proposal papers for future evaluations. In this way we hoped that we could provide a forum for exchange of information and to allow positive criticism of the past studies and the proposals. Authors were given 20mins of presentation time and another 20mins of discussion time, and a 50-minute session was allocated at the end purely for discussion. The presentations ranged from the very practical to the very theoretical and we were not short of a little controversy - one paper suggested that users could be simulated. In preparing their presentations the authors were encouraged to include discussion points, which in some cases led to some very interactive sessions.

What came out of the workshop was that there was a need to bring together resources for the user evaluation community in Information Access. A quick show of hands revealed that many of the attendees had not had any formal training in user evaluations. Many researchers interested in user evaluations were primarily from the Information Access field and had built a system which then required a user-based evaluation.

Fortunately there is a wealth of knowledge and experience in running user-based experiments from psychology and human-computer interaction. It is important that researchers are aware of the principles that have been developed in these fields. However every user

study is unique, and there will be a very substantial factor for user studies in information access that requires specific know-how and experience that must be nurtured and shared in the community. To address this the organisers will be looking to building a resource page for the community based around the workshop.

And finally, while there are many challenges to designing, performing, and analysing the results of a user study, there are also many benefits. One obvious benefit highlighted by many of the experienced researchers was the interaction with the users. People participating in the studies will always bring variety to your day, and you could never predict what they might do (or could you...?)

Alex Bailey leads the Document Analysis team at Canon Technology Europe. His interests focus on the use of document clustering, information extraction, and information retrieval for corporate document management systems. He can be contacted via: alexb@cte.canon-europe.com

Acknowledgements As organisers, Alex Bailey, Ian Ruthven, Leif Azzopardi would like to thank the program committee for their much-valued contributions and prompt reviews. We would also like to thank the authors and the attendees for providing the all-important content and of course, the discussion.

The CoLIS5 workshops were sponsored by the BCS Information Retrieval Specialist Group.
http://www.cis.strath.ac.uk/external/colis5/
http://www.cis.strath.ac.uk/~ir/evaluation.html

Forthcoming Events
Edited By Andy MacFarlane

AMR 2005: 3rd International Workshop on Adaptive Multimedia Retrieval
Glasgow, Scotland. 28-29 July 2005. Part of the IR Festival in Glasgow, & co-located with the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005) and INEX 2005 Workshop on Element Retrieval Methodology.

MRC2005: Modelling and Retrieval of Context
Edinburgh, Scotland. 30 July - 5 August 2005. A Two-day workshop to be held at IJCAI 2005, on context aware applications that are particularly relevant to IR. http://mrc2005.workshop.hm/

SIGIR'05: 28th ACM Conference on Information Retrieval
Salvador, Brazil. 15th - 19th August 2005. The premier world conference on research in Information Retrieval. Covers all major areas of information retrieval.
http://www.dcc.ufmg.br/eventos/sigir2005/.

Workshop on Link Discovery: Issues, Approaches and Applications (LinkKDD-2005).
Chicago, IL, USA. 21 August 2005. Part of The Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. A general workshop of link discovery issues with a theme on information retrieval.
http://www.isi.edu/LinkKDD-05/

IDDI-05: DEXA 2005 Workshop on Integrating Data Mining, Databases and Information Retrieval
Copenhagen, Denmark. 22nd August 2005. This workshop solicits papers on the issue of the problem of very large data sets looking at IR technologies and the problems researches in the field share with others who work in Data Mining and Databases. http://iddi05.unibg.it/

International Conference on Next generation Web Services Practices (NWeSP'05)
Seoul, Korea. 22-26 August 2005. A conference which addresses issues of new technologies as applied to the web, including the semantic web and ontologies. http://www.nwesp.org/

European Summer School in Information Retrieval (ESSIR 2005)
Dublin, Ireland. 5-9 September 2005. An event which introduces all the main concepts and 'hot topics' in information retrieval to those new to the area. http://www.cdvp.dcu.ie/ESSIR2005/

ISMIR 2005: 6th Annual Conference on Music Information Retrieval
London, England, UK. 11-15 September 2005. The first conferences series to be established on access to digital music materials.
http://ismir2005.ismir.net/

2nd International Workshop on Text-Based Information Retrieval (TIR-05)
Koblenz, Germany. 11-14 September 2005. A workshop held in conjunction with the 28th annual German Conference on Artificial Intelligence. Text retrieval with a focus on AI techniques.
http://www.aisearch.de/tir-05/

ECDL 2005: 9th European Conference on Research and Advanced Technology for Digital Libraries
Vienna, Austria. September 18-23, 2005. The major European conference on digital libraries, and associated technical, practical, and social issues, bringing together researchers, developers, content providers and users in the field. http://www.ecdl2005.org/

2005 IEEE/WIC/ACM Joint conference on Web Intelligence and Intelligent Agent Technology WI 2005/IAT 2005
Compiegne University of Technology, France. 19-22 September 2005. Includes a workshop on Open Source Web Information Retrieval.
http://www.comp.hkbu.edu.hk/WI05/

First International Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces
Reading, England, UK. 3-4 October 2005. A general workshop on personalisation that is applicable to information retrieval applications.
http://www.icete.org/workshop1.html

Human Language Technology Conference (HLT/EMNLP 2005)
Vancouver, B.C., Canada. 6-8 October 2005. A general conference on Natural Language Processing with a theme on information retrieval.
http://www.cs.utexas.edu/~ml/HLT-EMNLP05/

AIRS 2005: Second Asia Information Retrieval Symposium
Jeju Island, Korea. October 13-15, 2005. A symposium addressing all aspects of information retrieval from theories to user studies to applications. http://www.airs2005.org/

IADIS INTERNATIONAL CONFERENCE WWW/INTERNET 2005
Lisbon – Portugal. 19-22 October 2005. A general web conference with many themes related to information retrieval such as Digital Libraries and Semantic Web. http://www.iadis.org/icwi2005/

The 14th ACM Conference on Information and Knowledge Management (CIKM)
Bremen, Germany. 31 October – 5 November 2005. A general conference on knowledge management with themes on information retrieval, including a workshop on Geographical IR and peer to peer networks for IR. http://www.tzi.de/CIKM2005/

SPIRE'2005: String Processing and Information Retrieval
Buenos Aires, Argentina. 2-4 November 2005. A well regarded annual conference which focused on String Processing of all kinds including Information Retrieval. It has a strong South American focus.
http://www.la-web.org/spire2005

DocEng 2005: ACM Symposium on Document Engineering 2005
Bristol, England, UK. 2-4 November 2005. A symposium which is devoted to the dissemination of research on models, tools and processes that improve our ability to create, manage and maintain documents.
http://www.hpl.hp.com/conferences/DocEng2005/

2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies
London, England, UK. 30 November – 1 December 2005. Workshop with a theme on multimedia information retrieval, using ideas such as relevance feedback and ontologies.
http://www.acemedia.org/ewimt2005/index.html

ASIAN'05 Tenth Asian Computing Science Conference Data Management on the Web
Kunming, China. 7-9 December 2005. Focus of this particular conference is data management on the web, including issues such as search and information retrieval, Semantic web etc.
http://www.ynu.edu.cn/asian05/

Book Reviews

The following titles are currently available for IRSG members to review:

- Text Mining, ISBN: 0-387-95433-3
- Information Visualization, ISBN: 1-85233-789-3
- Information Retrieval, ISBN: 1-4020-3003-7
- Web Dynamics, ISBN: 3-540-40676-X
- Towards Knowledge Portals, ISBN: 1-4020-2053-8

To obtain a copy, all you need do is write a review for publication in Informer. Simply email us at irsg@bcs.org.uk with your contact details, including full postal address. This offer is available to IRSG members only, and is on a "first come first served" basis - so be prepared to be disappointed if you're not quick!

Industry Day

For the first time in its history, the IRSG's annual conference (ECIR) will be followed by a special day devoted to the interests and needs of IR practitioners. The Industry Day after ECIR 2006 is devoted to the challenges involved in designing and developing operational IR products and services, and aims to build bridges between IR specialists in industry and academia. This forum presents an opportunity for commercial organisations and individuals to share their work with a wider audience, and for researchers to learn more about the issues and problems faced by IR practitioners in developing practical solutions for information search & retrieval.

The scope of Industry Day 2006 covers all the areas addressed by the ECIR 2006 conference, but we are particularly interested in presentations and demonstrations of the following:

- search engines (web & enterprise)
- information architecture
- knowledge & content management
- data mining & visualisation

Industry Day 2006 will be held on April 13, at BCS HQ in central London (10 mins by Tube from the main ECIR conference venue). A separate one-day registration rate will be available.

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
- Yahoo 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 (http://www.bcs.org/bcs).
To automatically receive your own copy of Informer, simply join the IRSG via the BCS website (http://irsg.bcs.org/join_form.php).