

Contents

| | |
|---|----|
| Editorial | 1 |
| by Tony Rose | |
| Forthcoming Events | 2 |
| edited by Andy MacFarlane | |
| Feature Article | 4 |
| "A First Timer's Perspective on ECIR" | |
| by Terence Clifton | |
| Feature Article | 7 |
| "Phrases and Web Retrieval" | |
| by Gilad Mishne & Maarten de Rijke | |
| My PhD | 10 |
| "Distributed Knowledge Sharing in Mobile Settings" | |
| by Zia Syed | |
| Contact Information | 12 |



Welcome to the Spring edition of Informer. As many of you know, it's around this year that we usually have our annual conference (ECIR) – and this year was no exception. In fact, much of this issue is devoted to ECIR 2005: we have a conference review by Terence Clifton, and our feature article this time is an abridged version of the best paper, "Phrases and Web Retrieval", by Gilad Mishne and Maarten de Rijke. To complement this we'll be continuing our newly introduced "MyPhD" column, with an article on "Distributed Knowledge Sharing in Mobile Settings" by Zia Syed, and as usual Andy MacFarlane will be providing a timely roundup of IR related events around the world. So all in all, a very research-oriented issue this time! Despite everything I said about the research/practitioner balance in the last issue ...anyway, more on that below.

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.

In the meantime, those of you who attended ECIR will no doubt be aware that Ayse Goker will be standing down as IRSG Chair after 5 years of hard labour. We wish her all the best in her future work in the field. Consequently, we have a vacancy on the committee at present. As I mentioned last time, I think it would be valuable to the group as a whole to have more industrial representation on the committee, so if you are an IR practitioner, or have ideas about how we could promote our activities more widely to commercial organisations, please do get in touch.

One other development that you may have heard about at ECIR 2005 is that for next year's conference we plan to

introduce a new event, which we're tentatively calling Industry Day (see sidebar on p12). As you might have gathered, the aim of this event is to provide a forum devoted to the interests and needs of IR practitioners. This concept isn't completely new (in fact it's been blatantly plagiarised from the BCS HCI Group) - but it's relatively new to the IRSG. And, as with any new idea, there's a risk. But if we're serious about engaging with the practitioner community then I believe this is exactly the sort of event we should be encouraging. I hope you'll give it your support. In the meantime, we'd very much welcome your feedback on this idea (positive or otherwise).

And finally, thanks for all the valuable feedback on the last edition of Informer. It's encouraging to hear that it was so well received, but there are always things we can do to improve, so keep the reactions coming in. I'm also pleased to report that membership has increased by over 60% following ECIR - not bad for one event. Let's hope we can keep the momentum going. And don't forget - we don't have a monopoly on content production here! So if you have something topical you'd like to say, or an article you'd like to publish, by all means get in touch. We don't bite - in fact; we'll do our best to help you turn your idea into a successful article. Just drop us a line. Until the next issue,

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

Events

Edited By Andy MacFarlane

AVIVDiLib'05: 7th International Workshop on Audio-Visual Content & Information Visualization

Cortona, Italy, May 4-6, 2005. Focus on new solutions for content organization, access and interaction in Audiovisual digital libraries.

<http://delos.dis.uniroma1.it/Workshops/default.aspx>

PERVASIVE 2005: 3rd International Conference on Pervasive Computing

Munich, Germany, 8-13 May, 2005. Focus on ubiquitous, continuous, and self-organized technologies in IT applications which addresses some of the major issues in IR.

<http://www.pervasive.ifi.lmu.de/>

ISI-2005 IEEE International Conference: Intelligence and Security Informatics

Atlanta, Georgia, U.S.A, 19th - 20th May 2005. Focuses on security issues as pertaining to the intelligence community. Of particular interest to the IR community is the issue of Information Sharing and Data Mining theme of the conference, particularly the use of Digital Libraries as applied to archiving, preserving and searching intelligence information.

<http://ecom.arizona.edu/ISI/>.

Multimedia and the Semantic Web

Heraklion, Crete 29th May 2005. One day workshop which aims to bring together researchers and practitioners in the multimedia and Semantic Web domains in order to assist in forming bridges between the communities for mutual benefit.

http://www.acemedia.org/ESWC2005_MSW/

COLIS5: 5th International Conference on Conceptions of Library and Information Science

Glasgow, UK, June 4-9 2005. Focus on exploring and understanding information-centred disciplines. It will examine how the issue of context influences the design and use of information access technologies.

<http://www.cis.strath.ac.uk/external/colis5/>

I-Know'05: 5th International Conference on Knowledge Management

Graz, Austria, 29th June - 1st July 2005. The focus is on application-oriented research for Knowledge Management. The conference has a number of tracks, which include: Knowledge and Information Visualization (KIV'05), Knowledge Discovery and Semantic Technologies (KDaST'05) and Knowledge Sharing in Research and Higher Education (KSR'05).

http://www.i-know.tugraz.at/conference/i-know05/iknow05_home.htm

PIA 2005

Edinburgh, July 25, 2005. Workshop on New Technologies for Personalized Information Access at User Modelling 2005. Addresses the issue of using personalisation to improve the process of information retrieval.

<http://irgroup.cs.uni-magdeburg.de/pia2005/>

AMR 2005: 3rd International Workshop on Adaptive Multimedia Retrieval

Glasgow, 28-29 July 2005. Deals with the issue of access to multimedia information on one hand and indexing this data on the other. This workshop is part of the Glasgow IR Festival (<http://www.dcs.gla.ac.uk/irfest/>), and it is co-located with the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005) in Edinburgh.

<http://www.dcs.gla.ac.uk/amr2005/>

MRC2005: Modelling and Retrieval of Context

Edinburgh, 30 July - 5 August 2005. Workshop held as part of IJCAI 2005. Covers context-aware applications for IT that are particularly relevant to IR. The goal is to bring members of different communities together to discuss their problems.

<http://mrc2005.workshop.hm/>

Learning in Web Search

Bonn, Germany, August 7, 2005. Workshop held as part of the 22nd International Conference on Machine Learning (ICML 2005). Addresses the issue of machine learning as applied to web search in both theory and applications.

<http://cosco.hiit.fi/search/learninginsearch05/>

Methods and Applications of Semantic Indexing

Copenhagen, Denmark, Tuesday 16th August 2005. Workshop at the 7th International Conference on Terminology and Knowledge Engineering. Focuses on the semantic web, in particular the gap between classic information retrieval and terminological methods in order to improve semantic indexing.

http://wortschatz.uni-leipzig.de/~fwitschel/WS_SemanticIndexing.htm

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/>

IDDI-05: DEXA 2005 Workshop on Integrating Data Mining, Databases and Information Retrieval

Copenhagen, Denmark, August 22, 2005. This workshop solicits papers on the issue of the problem of very large data sets looking at IR technologies and the problems researchers in the field share with others who work in Data Mining and Databases.

<http://iddi05.unibg.it>

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/>

AIRS 2005: Second Asia Information Retrieval Symposium

Jeju Island, Korea, October 13-15 2005. Covers all aspects of information retrieval from theories to user studies to applications.

<http://www.airs2005.org>

CLIHC'05: Interaction for Inclusion, Cuernavaca, Mexico, October 23-26, 2005. Of particular interest to IR researchers and practitioners working in the area of HCI.

Covers interactive systems development, methodologies and tools, evaluation reports, theories and models.

<http://www.cihc2005.org/>

DocEng 2005: ACM Symposium on Document Engineering

Bristol, UK, November 2-4, 2005. Focuses on models, tools and processes that improve our ability to create, manage and maintain documents. Deadline for papers is 6th May 2005.

<http://www.hpl.hp.com/conferences/DocEng2005/>

SPIRE'2005: String Processing and Information Retrieval

Buenos Aires, Argentina, November 2-4, 2005. Conference focused on String Processing of all kinds including Information Retrieval. It has a strong South American focus. Deadline for submissions is 7th June 2005.

<http://www.la-web.org/spire2005>

IR Festival Glasgow, UK July 24-30 2005

A full week of IR activities hosted by the University of Glasgow, including:

- IR & Theory Workshop
- ESF Exploratory Workshop on IR in Context
- Adaptive Multimedia Retrieval & MMKM (Multimedia Knowledge Management Network) Workshop
- INEX 2005 Workshop on Element Retrieval Methodology

Registration for all events is *free*. However, there are only a limited number of places. See <http://www.dcs.gla.ac.uk/irfest/> for further details.

Feature Article:

"A First Timer's Perspective on ECIR"

By Terence Clifton



The annual European Conference on Information Retrieval, now in its 27th year, was held in Northern Spanish city of Santiago de Compostela, and hosted by the "Escuela Tecnica Superior de Enxeñería (ETSE)" (Technical School of

Engineering) of the University of Santiago de Compostela (<http://www.usc.es/>). As a first-time attendee of ECIR, I'd like to take this opportunity to share my experience of ECIR05 with you, and give you a flavour of what you may have missed if you were unable to attend ECIR this year. I am a relative newcomer to the Information Retrieval field in general, with my doctoral studies actually focussed in the field of Computer Graphics. I was introduced to IR research as a result of being involved in the development of a Question Answering system within my department at the Bangor University, and have since taken an active interest in the field, both at the specific QA and more general IR levels. As a result of this, I'm hoping that you'll find that this article provides an interesting perspective on both the conference itself, and the technical content, although I will apologise in advance for any misconceptions or errors that occur as a result of my naivety in the field.

About the Location and Venue

Situated in the North West of Spain in the region known as Galicia, Santiago de Compostela is famous for its religious influences and imposing architecture. Although not a typical tourist destination it is the final destination on the medieval route of pilgrimage Camino de Santiago (Way of Saint James), which attracts large numbers of people throughout the year, and as a result the city is well connected in terms of transport links.

The conference venue itself was very well suited to the event, providing an excellent lecture theatre for the main presentations, with adjoining breakout and cafeteria facilities. Plasma screens were situated around the facility detailing the upcoming program, and showing snapshots taken from the previous day's events, which was a nice touch by the organisers, and added the wholly professional presentation. Meals were provided in the University restaurant, and contrary to most conferences I've previously attended (all non-IR conferences it has to be said), the food was of a good standard, with a wide selection of options and most dietary tastes catered for.

The Technical Programme

As a relative newcomer to the field, I was pleasantly surprised by the breadth of papers presented at ECIR05 (I was expecting everything to be about web-search and its various incarnations), all of which were of a good technical standard. The keynote addresses, provided by Keith van Rijsbergen (University of Glasgow) and Ricardo Baeza Yates (ICREA-Universitat Pompeu Fabra, Spain, and University of Chile) were stimulating, and provoked some interesting discussion, both within the question answer session that immediately followed, and through the course of the conference.

As a general rule, each session of the conference contained a good selection of related papers and attracted an interested and knowledgeable audience. As a student myself, it was particularly good to see that many of the questions and comments which followed individual presentations were of a constructive nature, and I'm sure would be useful to all of us aspiring academic researchers.

Of the wide variety of interesting papers presented, I found a number particularly stimulating. The work presented by Evelyn Balfe from University College Dublin on collaborative web search (as part of the bigger ISpy project) (<http://ispy.ucd.ie>) was very interesting, and offers an interesting new direction for improving search effectiveness in community-biased situations. The presentation by Henrik Nottelmann on the PIRE system (http://www.is.informatik.uni-duisburg.de/bib/docs/Nottelmann_05.html.en) showed promise as an open system on which

to prototype IR developments, with the necessary fundamental tools provided by the probabilistic Datalog environment offering a number of benefits over bespoke developments. Somewhat surprisingly I was also impressed by the presentation from Kamal Ali from Yahoo Inc (<http://www.yahoo.com>) on cost effective approaches to human relevance evaluation for search engines. I was expecting a somewhat superior attitude, and limited disclosure due to the corporate nature of the presenter, but was pleasantly surprised by how candid he was about the work being carried out at Yahoo, and how the research he presented fits in with that. Relevance feedback is a major and difficult issue in search engine effectiveness, and although the presentation did not present any kind of silver bullet, it did concisely summarise the current situation, and provide a number of stimulating avenues that could be explored in future research.

In many of the previous conferences I've been to, it's often been the case that the best student paper award goes to the student with the most well known supervisor (maybe that's just me being cynical, but the result, in my opinion rarely reflects the quality of presented papers at the conference). I'm pleased to say that ECIR05 bucked that trend by giving the best student paper award to Gilad Mishne (<http://staff.science.uva.nl/~gilad/>) from the University of Amsterdam, for his paper entitled "Boosting Web Retrieval through Query Operations". The paper addressed a key area of information retrieval, provided a good set of foundation results, and was well written. The presenter did an excellent job in front of the audience, with a humorous, yet technically effective presentation of his work, and was more than competent in fielding questions from the participants.

Having discussed some of the papers that interested me within the technical programme at ECIR05 it just leaves me to reflect on how my own presentation went. Not being from the IR area, I was a little nervous about how the work would be received, but the audience were attentive and provided some useful comments and criticisms (constructive of course) to guide me in taking the work further. The chance to present at ECIR has given me a greater insight into the way IR research is both carried out and presented, and I must take

this opportunity to thank the ECIR committee and particularly CEPIS (<http://www.cepis.org>) for providing me with a substantial travel grant which made my attendance and participation at ECIR05 much easier.

The Social Programme

My preconceptions of the types of people who attend Information Retrieval conferences were, I'm afraid to say, somewhat stereotypical, and I went to Santiago with the expectation of stuffy conversations about precision and recall, click through mining, ontological metaphors and query optimisation. Needless to say I was mistaken, and in fact the ECIR attendees were very welcoming and sociable. The first evening of the conference was allocated to poster presentations and a drinks reception, and the copious amounts of wine made for a friendly and relaxed atmosphere

The posters presented were of a high standard, and covered some very interesting and relevant topics. I was particularly impressed by the work of the Glasgow team related to the Terrier (Terabyte Retriever) search platform (<http://ir.dcs.gla.ac.uk/terrier/>), presented at the conference by Craig McDonald. Also, despite my attempts, my graphical background came through as I found myself very interested in the SnapToTell system (http://www-mrim.imag.fr/publications/2005/CHE05/chevall-et05a_ECIR05_SnapToTell.pdf) presented by Jean-Pierre Chevallet from the IPAL-I2R Laboratory, Singapore. The system provides ubiquitous access to directory information through mobile network location, GPS, and image matching techniques, and was demonstrated as a guide to tourists, who could send an image to the SnapToTell Server using their camera phone, and receive relevant information as a response.

Overall the atmosphere, the friendliness of the other participants and attendees, and the excellent social programme provided by the organisers was a great success, and a credit to the ECIR and wider IR community.

Final Thoughts

I've been to a number of graphics-related conferences in my short period involved in academic research, but ECIR was my first

major venture into the world of Information Retrieval. I, obviously, had a number of preconceptions, both about the reception I could expect as a presenter and the type of researcher I would encounter within the field. Needless to say, many of these were dispelled quickly by the friendly, approachable people I met at ECIR05.

The conference was extremely well planned and organised, with excellent facilities provided by the University of Santiago de Compostela, and well-prepared, friendly, and knowledgeable conference organisers and session chairs. The technical papers were of a good standard, and covered a wide variety of strands within the IR field, providing something for everyone, and the social interaction, and opportunities to meet and collaborate with like-minded researchers was far better than I had expected. Overall I can't praise the organisation of the conference highly enough, and the co-chairs and team of local organisers should be proud of a job well done.

As a first time visitor to ECIR I can safely say that I enjoyed the conference, both from a technical and social perspective, and will definitely be on the registration lists for future events – see you in London 2006.

Terence Clifton is a PhD research student in the School of Informatics at the University of Wales, Bangor, where, despite his interest and involvement in Information Retrieval and Artificial Intelligence, he is actually studying for a doctorate in the field of Computer Graphics. He is however, and integral member of the Artificial Intelligence and Intelligent Agents research group at Bangor, and one of the main developers on the groups agent-based Question Answering System – QITEKAT. He can be contacted via:
terence@informatics.bangor.ac.uk

Feature Article:***"Phrases and Web Retrieval"***
By Gilad Mishne & Maarten de Rijke

It is generally accepted that stating information needs in a more focused way leads to more precise results. For example, for a tourist arriving with her dog to the Netherlands, the query pet-friendly hotels in Amsterdam will

probably yield more accurate results than a search for hotels in Amsterdam. This is why intuition leads us to believe that the usage of phrases when expressing information needs will improve the results of the search: while a search for white house pictures may return both pictures of the residence of the U.S. president and pictures of various houses colored in white, the more focused search for "white house" pictures will narrow the results to the presidential residency by guaranteeing that the words "white" and "house" appear consecutively, in that order, in the retrieved information. Now, it can be argued that this improved performance comes at the cost of missing some truly relevant results (in technical terms, increases precision at the price of decreased recall); however, when dealing with large quantities of information such as those existing on the World Wide Web, this is not an important factor. Studies show that users are concerned mainly with the accuracy of the first 10 or 5 results displayed: they wish to get "some relevant data" ranked highly, and not necessarily "all relevant data."

Following this intuition, the use of phrases has been researched extensively in the last 30 years in the field of Information Retrieval. Sadly, the general conclusion reached was that given a good ranking formula, the use of phrases has no substantial effect on performance. This has been validated a large number of times, using various experimental settings and measurement metrics. However, an important characteristic of all these experimental environments was that the collection of documents on which the search was performed was basically plain-text

documents, usually newswire feeds of a number of years.

In this article, we revisit the usage of phrases (and additional, less restrictive multi-word expression units) in retrieval. This time, however, we focus on web retrieval: searching in HTML documents. The questions we set ourselves are:

- Is the use of phrases beneficial in the web setting?
- If so, why is it different from plain-text documents?
- Is there a simple, robust way of automatically using phrases to improve search?

Web documents and phrases

Why do we hypothesize that web documents may exhibit a different behaviour with phrase searches than plain-text documents? To answer this, we point out a prominent feature of HTML pages, often used in the web retrieval setting. HTML was originally devised to describe the physical layout of a page, marking entities such as page title, emphasized text, font size and so on. This markup has been successfully used in numerous ways to improve search performance from web pages, for example by assigning a higher weight to headline text than to paragraph text; this technique is sometimes referred to as multiple document representation, meaning that the document is separated to different "representations" (or fields) — the title, the headline text, the anchor text (text present in all links in pages pointing to a certain page), and so on. Different fields are then searched separately, and results are combined to form a single ranked document list, assigning more importance to representations such as headline text.

Examining these different fields, we see that unlike plain-text documents, some of them are highly rich in phrases: such are the page title, the anchor text, and even the URL text. This is a direct result of the purpose of these fields: the page title is a short, usually human generated summary of the topic of a page, and in many cases is or contains phrases. Similarly, the anchor text and URL text are very short descriptions of the page, meant to summarize its content to a single soundbyte.

Additional short, descriptive fields are the META "keywords" and "description" sometimes associated with web pages, and containing a short list of important words and phrases relevant to the page.

In addition to the difference in the searched documents, in the case of web searches the queries themselves are also rich in phrases. The informational web queries used in evaluations such as TREC contain a very high percentage of phrases (almost 80% of the queries with more than one word are or contain phrases); in the case of web query logs from actual commercial search engines, phrases are also found in the majority of the queries (although percentages are slightly lower).

The combination of the abundance of phrases in certain fields of web documents, and their frequent usage in queries on the web leads us to hypothesize that the use of phrases for searching the web will result in a higher gain to retrieval performance than in the case of plain-text documents that has been explored extensively in the literature.

Detecting phrases and expanding queries

There are three general approaches to phrase recognition in text: *syntactical*, *statistical*, and *lexical*. Syntactical approaches rely on linguistic analysis of the text, using features such as part-of-speech tags or dependency relations to identify words which constitute a phrase. Statistical approaches rely on *collocation* information: phrases are identified by looking at the entire corpus and examining which words tend to appear together. Finally, lexical approaches are based on pre-constructed lists of typical phrases in a language. Naturally, hybrid methods exist which combine ideas from the three techniques.

Given the average length of web queries, and the fact that they are mostly ungrammatical, using the syntactical approach is unlikely to yield good results. Statistical approaches generally provide good results, but need to be adjusted per corpus separately: words that tend to appear together in one collection might not do so in another. Finally, lexical approaches require obtaining large amounts of phrases in advance and are not scalable. We

therefore take a simpler, perhaps naïve approach to phrase detection: we consider *every* combination of consecutive words from the query (of any length) as a phrase. For example, given the user-supplied query *well water contamination*, we identify all of the following as phrases: "well water", "water contamination," and "well water contamination." All phrases detected this way are then added to the original query as phrase terms, creating a long query containing the original terms as well as the more focused phrases. With modern retrieval ranking formulas, this results in rewarding documents which contain as many phrases as possible from the query (and still matching documents which contain the separate words, but not the phrases).

The reasoning behind this seemingly shallow approach is as follows:

- Empirically, this approach captures the vast majority of phrases in the queries.
- While this method also makes mistakes, they are unlikely to affect the search performance. For example, assume that we are given the query *automobile emissions vehicle pollution*; while we successfully capture the phrases "automobile emissions" and "vehicle pollution," we incorrectly identify non-phrases such as "emissions vehicle." However, since they are not phrases, it is very unlikely that they appear in the document collection, so including them in a query will not change the retrieval results (in most modern, non-boolean retrieval formulas).
- In fact, the identification of non-phrases might even improve results by matching texts which contain the phrase words in high proximity, if not as a real phrase. Returning to the "emissions vehicle" example, this "phrase" will match (given the standard stemming and stopping processes common in web search) multi-word units such as "emitted from a vehicle" or "emissions of vehicles".
- Finally, this light-weight approach is extremely fast and robust, relying on no external algorithms such as linguistic analyses and statistics.

In addition to the expansion of queries with phrases, we take a similar approach to expand queries with *proximity terms*. These are also multi-word terms; the difference between them and phrase terms is that they are permitted to match not only documents which contain the phrase “as-is”, but any document which contains all the words in the term inside a window of K words (K is a parameter, values are 5-15).

Experiments

To test our hypothesis, we used 125 informational queries from the web retrieval evaluation at TREC 2003 and 2004; the average length of the queries was 2.4 words (which is the same length as reported by major commercial search engines). We chose informational queries rather than navigational queries both because the performance on the latter is very high already, and does not require additional substantial work, and because phrases that appear in navigational queries tend to be proper names; as such, they are unlikely to appear in any form except as phrases, and converting them to phrases will not change the retrieval performance. The document collection we used was the TREC .GOV collection — a crawl of the .gov domain consisting of 18GB of text in 1.25 million HTML documents.

We tested our phrase-expansion method on two versions of the collection: in the first version, documents were treated as plain-text documents, meaning that all the fields (title, anchor text, etc.) were combined into a single representation of the document. In the second version, we separated the content in the documents according to the following fields: title, URL, body, and anchor text. On the first version of the collection we observed, like many others before us, no apparent gain from using phrases: sometimes a small increase in performance, sometimes a small decrease. However, on the second version we observed substantial improvements of up to 23% — depending on various parameters and the metric measured. The expansion was especially helpful for queries of length 2–3, which constitute the vast majority of all informational queries.

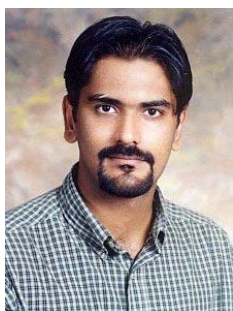
Conclusions

We set out to examine the usage of phrases in the domain of web retrieval. Although past research on retrieving phrases has generally not shown substantial improvements, we hypothesized that for HTML documents this may be different. The existence of short, phrase-rich, highly descriptive fields (such as title and anchor text) in these documents suggested that using phrases for their retrieval can provide a better outcome than using phrases for plain-text retrieval. Our experiments support this hypothesis, showing that the usage of phrase expansion — even with a very simple phrase detection method — substantially improves retrieval effectiveness for web documents.

Gilad Mishne received his undergraduate degree from the Technion (Israel Institute of Technology), and after a number of years in industry returned to the academic world to get an M.Sc. from the University of Amsterdam, where he is currently pursuing his Ph.D. Gilad's research is focused on Information Retrieval in the emerging domain of blogs. He can be contacted at gilad@science.uva.nl.

Acknowledgements: This paper is a summarized version of “Boosting Web Retrieval with Query Operators”, appearing in: D.E. Losada and J.M. Fernández-Luna, editors, *Advances in Information Retrieval: Proceedings 27th European Conference on IR Research (ECIR 2005)*, LNCS 3408, Springer, pages 502-516, 2005.

Both authors were supported by the Netherlands Organization for Scientific Research (NWO) under project number 220-80-001. Maarten de Rijke was also supported by NWO under project numbers 365-20-005, 612.069.006, 612.000.106, 612.000.207, 612.066.302, 264-70-050, and 017.001.190.

My PhD:***"From P2P to U2U: Distributed Knowledge Sharing in Mobile Settings"******By Zia Syed*****Introduction**

With the emergence of wireless computing environments, information users are increasingly becoming mobile in a social world. The information needs of the mobile users are becoming more demanding yet constrained with numerous limiting factors such as interaction time, wireless communication instability etc. In mobile computing environments, users carrying portable computing devices have access to a shared infrastructure independent of their physical location. This provides a flexible communication model between people and networked services. At present, the current research trend in mobile systems focuses upon the communication aspects of technology. However, communication technology is subsequently being used to fulfil people's information needs. In a pervasive computing environment, information is distributed across numerous stationary as well as mobile information nodes.

Scenario

For the illustration purpose, consider the scenario; you have just arrived in an unfamiliar city planning to do your shopping. Faced with a multitude of gift shops and with no idea where to start, you use your handheld to seek out other shoppers' opinions about the best places to shop. Your device receives relevant opinions and advice by people within your physical proximity. Using this personal knowledge as your guide, you set off on your shopping trip. Although this scenario may sound slightly futuristic, with the recent advances in wireless communication technologies, particularly in Personal Area Networking (PAN) [<http://www.comp.lancs.ac.uk/~kortuem/publications/p2p2001.pdf>], coupled with the ever-

increasing rate of mobile device adoption around the globe, we see this scenario becoming a reality, sooner rather than later. This scenario setting is typically attributed as a Mobile Social Cyberspace [http://www.smartmobs.com/book/toc_7.html]

Pervasive Information Retrieval

With the evolution of the pervasive computing infrastructure, more and more objects in our everyday environment are getting equipped with tiny processors, wireless communication modules, and embedded software, making them capable of interacting and communicating with other artefacts. As these devices communicate, they form tiny yet content-rich mobile adhoc information-sharing networks (MANET). In a MANET, information spreads across each node (peer or digital artefact) on the network and the overall configuration of the system is highly dynamic. Therefore, it is important to adopt approaches that respect the inherent symmetry in the network of wireless devices; the peer-to-peer (P2P) information-sharing model naturally matches this situation [<http://jabber.comp.rgu.ac.uk/research/publication.php?id=897>]

Social CyberSpaces

With the penetration of Internet technologies, people have moved out of confined local communities and works groups, and now have involved themselves into far-flung internet-enhanced social relationships with friends, families, workmates, neighbours or simply with peers of similar interests. The evolution of social cyberspaces has put the developed world in the midst of a paradigm shift in terms of the ways people, organisations, and social groups are now connected to one another. Therefore, the success of future pervasive computing environments lies in the prolific implementation of technologies that help spawn Group Forming Networks (GFN) [<http://www.reed.com/Papers/GFN/reedslaw.html>].

Research Aim

The aim of this PhD research is to tackle the problem of distributed information retrieval in P2P networks, particularly in mobile adhoc settings. It also aims to investigate and understand the social aspects of mobile

cyberspaces, which can be used to design a founding framework for developing a range of software applications. Applications that would help evolve group-forming networks in mobile adhoc environments.

In support of this research aim, a modular software framework, named IRMAN (Information Retrieval system for Mobile Adhoc Networks)

[<http://www.springerlink.com/index/RJ0CY15RGLGXFVH5>] has been designed. The IRMAN framework enables the development of software applications for mobile environments. It incorporates an information-processing engine that uses context-awareness (particularly user and device context) for effective information retrieval in mobile adhoc environments

[<http://jabber.comp.rgu.ac.uk/research/publication.php?id=72>]. It also provides application layer connectivity between wireless devices in a P2P fashion, and enables content searching and sharing in an adhoc manner. The modular design of the framework supports component based development of a variety of P2P applications (e.g. File swapping, Instant Messaging etc.) for handheld devices irrespective of communication, information processing and presentation technologies.

Conclusion

The vast manifold of wireless computing artefacts has started to populate a pervasive computing landscape, characterised by paradigms like context-awareness, spontaneous interaction, and adhoc networking. This research tackles the challenges of effective information retrieval in mobile computing environments. It aims to exploit the potential of the Peer-to-Peer communication model for an efficient user-to-user information retrieval experience in a mobile social world.

Zia Syed is a Computer Engineering graduate (2000) from NED University, Karachi, Pakistan; he obtained a Masters in Software Engineering from University of York, UK, in 2002. He is currently pursuing a Doctorate in Computer Sciences at The Robert Gordon University in Aberdeen, UK, under the supervision of Dr. Ayse Goker. There, his main research interests are in Pervasive Computing, and

Information Retrieval. He can be contacted at zs@comp.rgu.ac.uk

CoLIS 2005 WORKSHOP – Evaluating User Studies in Information Access

Wednesday 8th June 2005

BCS IRSG SPONSORED EVENT *Members get £5 pounds off!*

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. However, 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. 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. This workshop aims to:

- assess current user based studies in Information Access; discussing the advantages and disadvantages of their methodology,
- to provide a forum of discussion for proposed user studies submitted to the workshop, and
- to draw up a list of guidelines for future user studies.

For the workshop we solicit two types of contribution: proposed user studies and position papers on user evaluation. Submissions will be accepted up until the 1st of May.

For further information, contact Alex Bailey (alexb@cre.canon.co.uk) or visit the workshop website:

<http://www.cis.strath.ac.uk/~ir/evaluation.html>

Industry Day

For the first time in its history, 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 and 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 & navigation
- 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.

Authors

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.

If you have an idea for an article, please contact the editors at: irsg@bcs.org.uk.

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).