Informer

BCS
INFORMATION RETRIEVAL

In This Issue

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.

If you attended Search Solutions a few weeks ago then you'll no doubt appreciate how this event is gradually starting to carve out its own unique identity – an informal, interactive alternative to the highly choreographed, glossy brochure, convention hall style of event that otherwise seems to dominate the search industry. Or, to put it another way, we're small because as organisers we're not much good at marketing… but maybe that's a blessing in disguise.

Either way, if you weren't fortunate enough to be there, then Keely Flint's review on p2 will provide you with a topical summary of what you missed. This is followed by an abridged version of my own presentation "UI Design Patterns for Search & Information Discovery". Note that we're hoping to make all the presentations available from the event website soon – be sure to check back on: http://irsg.bcs.org/.

Another positive outcome of the day was the AGM – by co-locating with Search Solutions (and its generous post-event drinks reception) we were able to gain several new members, a number of which are now serving on the committee. I hope you'll join me in making them feel welcome.

Finally, don't forget that we still have a number of books available for review – check out the IRSG group pages on LinkedIn for up to date details. Until next time,

Best regards,
Tony

Tony Russell-Rose, PhD MBCS CITP
Editor, Informer
Vice chair, IRSG
Email: irsg@bcs.org.uk
http://www.information-interaction.net

Conference Review:

Search Solutions 2009

by Keely Flint

Search Solutions 2009 brought together insights from across the search community to an interactive day spent imagining 'what could information seekers search experience be like in the future?'

Throughout the packed agenda, the intellectual food for thought was first class. Expert speakers representing giants such as Google and Yahoo, joined Enterprise Search vendors and protagonists of open source search, in sharing their latest developments.

The day kicked off with a review by Mike Taylor of some of the techniques for learning ranking functions investigated at Microsoft research. An impressive Vivian Lin Dufour then introduced Yahoo! Search Pad, one of the innovative applications added to the new Yahoo! Search framework. Search Pad is a note-taking application that automatically tracks and organizes sites you find on Yahoo! Search to help you complete important tasks such as buying a car or planning a trip. This is a clever product development by Yahoo!, thankfully more user than ranking function centred and which ambitiously tries to move us beyond the typical ten blue links search experience.

The Google AdWords program was explained by Richard Russell as a search engine where ranking is determined by commercial considerations. Whilst AdSense is a generous source of revenue for Google and advertisers, Richard also stressed the importance to Google of pleasing the user, by ensuring that ads served are contextually relevant. Richard explained how careful Google are in using the information they have about search history and user intent. This is reassuring to hear as privacy and data ownership related to our online activity are a growing cause for concern.

Before lunch, Dave Mountain of Placr Ltd evangelised on location based services and applications. Dave emphasised that some technical and service challenges remain in the way of location based services reaching their potential, however if any innovation talked about today could bring about a total shift in our experience of search it would have to be this. His prediction for the future: "a world where **you** are the search engine".

Alan Oliver of Ex Libris, gave a thought provoking speech about librarianship in the digital age, how it has changed and how it needs to keep changing. Not long after, we were treated to a very brave Andrew Maisey of Unified Solutions, defending the benefits of taxonomy in content management. It's good to hear someone so confidently outspoken in the value of more formal classification structures, given that taxonomy, like librarianship and probably no thanks to the Dewey decimal system, has fallen out of fashion in recent years. However, Andrew hit it home that despite the usefulness of social tagging in guerrilla classifying the fast growing mass of online content, for those working in the gritty job of enterprise content management, taxonomies remain terribly important to ensure effective information retrieval.

'what could information seekers search experience be like in the future?'

No question at all of the value of a well classified corpus of enterprise content, but is this just a utopian dream? The real issue for the future remains that the cost burden of building and applying taxonomy manually to content, which still seems to be the only way to achieve high quality results, is the main reason that many organisations don't want to try.

Tony Russell-Rose shared the goals of Endeca to create a library of UI patterns for search and information discovery that they intend to share openly. As an Information Architect I really appreciated the good sense of this, knowing that not only good standards but

common standards in user experience help users to find information quicker and easier. It's incredible to see how many websites deploy search with a user interface that would alienate the bravest of users, no matter how good the underlying technology.

Tony explained that an enhanced search user experience is critical because search works when you know exactly what you want, but not when you don't know what you are looking for, or have the terminology to articulate it. A really good example of this is health information seekers behaviour, a context I'm very familiar with. Health information is one of the primary reasons users search online and typically, health information seekers are new to a subject and have no understanding of the terminology they need to research a subject. Tony also emphasised how context is paramount in information seeking behaviours – again I can relate to this in that health information seekers are often in a vulnerable state and impatient to find the information they need. It is perfectly evident how the heavily guided search and browse environment of the NHS Choices website, informed by a thorough user centred design process, reflects the unique needs of the nation's health information seekers.

After lunch, Greg Grefenstette of Exalead, talked about the maturation of search and search based applications. Greg showed us a future in which the gap between database applications and search is closing. Search based applications such as logistics tracking and customer management will invoke complex search across datasets and agile applications will change the way that information looks and is experienced by users. A bold vision from Exalead here.

Surprisingly and much to my relief, not one speaker mentioned Web 2.0 or social media at this event, however participatory culture at its best was demonstrated during the very open and interactive discussions. I enjoyed thought provoking conversation during the breaks with a Phd from the Open Universities Knowledge Media Institute researching the Semantic Web and another from UCL, an expert in Folksonomy and online tagging communities. So apart from the knowledgeable hosts and speakers that put together this event, the

interesting audience it attracted made it an event even more worth attending.

Keely Flint is currently an Executive MBA student at London Business School. Her experience gained through 10 years developing websites in the digital publishing and healthcare industries include: enterprise information architecture; semantics management and search.

Email: kflint.semba2010@london.edu
Linkedin: http://www.linkedin.com/pub/keely-flint/14/772/536

Informer

Feature Article:
UI Design Patterns for Search & Information Discovery

By Tony Russell-Rose

Design Patterns are principled solutions to common design problems in a particular field of expertise. Although originally developed and applied within the field of architecture, design patterns have since been applied to software development and more latterly, interaction and user experience design. This article examines the role of patterns in designing information search and discovery applications, describes some of the challenges involved in creating the Endeca UI Design Pattern Library (UIDPL), and explores some of the issues involved in maintaining and growing a pattern library as a resource for the user experience design community.

The Endeca UI Design Pattern Library

The Endeca UIDPL currently consists of approximately 20 patterns focusing on UI challenges unique to faceted navigation and information discovery. The library extends the work of others such as the Yahoo! Developer Network , Jenifer Tidwell, Martijn van Welie and Peter Morville. The patterns cover a wide variety of user experience (UX) elements, ranging from simple interface components such as breadcrumbs to more complex information structures such as search interfaces, search results pages, faceted navigation menus, and discovery tools such as comparative views and product 'configurators' (i.e. tools to guide the user through a process of configuring a product or service). An excerpt from the pattern for the "breadbox" (an interactive, breadcrumb-like component used to represent the user's navigational state) is illustrated in Figure 1.

Each entry in the pattern library is structured as follows:

- **Pattern name**: a concise label for the proposed solution;

- **Problem summary**: a brief description of the issue or problem that the pattern seeks to address;
- **Usages**: a list of the contexts in which the pattern is typically applied;
- **Constraints and challenges**: the primary issues that need to be considered when implementing the pattern;
- **Solution elements**: the design guidance that enables a UI designer to understand and apply the pattern (illustrated with visual examples wherever possible);
- **Anti-patterns**: examples of implementations that seek to address the same problem, but are non-compliant with the design guidance (i.e. one or more solution elements are violated).

Each design pattern may also include references to related design patterns and additional examples.

Creating a pattern library

Ideally, each design pattern should be backed up by an evidence base derived from rigorous primary research and empirical investigation. However, in a commercial setting it is not always possible to perform exhaustive investigation into the rationale for every solution element of every pattern. In addition, the pace of UI innovation and change can also require rapid publication and dissemination of so that best practice is shared at the earliest available opportunity. In practice, therefore, the patterns are based on evidence from a variety of sources, such as:

- Primary research and user testing, performed both internally and by the external academic community;
- Analytical models of human information seeking and discovery behaviour;
- Practical lessons learned from previous projects and implementations.

The design guidance provided by the patterns is presented at a level of abstraction that is applicable across a broad range of contexts. For example, a solution element such as "Provide users with the ability to sort the query results" could be implemented using clickable

Informer

Figure 1: The UIDPL entry for the "Breadbox"

table headers, a drop-down list, or radio buttons, etc. The choice of which particular control to use is then made by the UI designer based on their knowledge of the specific user context. Where possible, guidance includes criteria by which to select among such options, e.g. when to favour radio buttons over a drop-down list, etc.

Maintaining & growing a pattern library

Establishing a library requires a significant investment of time by key individuals with the promise of value delivered for the company, customers, and end-users. Maintaining and growing the library requires an ongoing commitment to document and proactively share lessons learned and best practices, in a fast paced and often pressured services and software development environment.

The pattern library itself is built as a wiki, with write access granted to members of the UX team. Others in the company have read access and the ability to provide comments and

questions directly in the Wiki. To maximise adoption, we openly encourage others to submit requests for candidate patterns, and respond to UI design queries by referring wherever possible to the Endeca UIDPL.

> ### "Establishing a pattern library requires a significant investment of time by key individuals"

At present, the Endeca UIDPL is currently only available internally and as such represents a proprietary resource whose value is realised when applied within customer projects and product development. In this case, the ROI for creating the UIDPL is both direct and measurable. But maintaining the UIDPL as a proprietary resource is just one of many ways to realise value from this investment. A contrasting approach is to make the UIDPL an open resource, accessible to public scrutiny and feedback. In this scenario, the ROI is perhaps less directly measurable, but the

Informer

benefits are enjoyed more broadly by the wider community and the feedback received enhances and extends the scope and value of the UIDPL itself. This issue merits further exploration within the company and more widely, but our current aspiration is to open the UIDPL to the broader community to accelerate exchange, growth, and innovation.

Tony Russell-Rose is currently Manager of User Experience at Endeca, an enterprise software company specialising in innovative solutions for information search and discovery. Before joining Endeca he founded UXLabs, a user experience consultancy specialising in technology innovation and applied R&D. Prior to this he was technical lead at Reuters, specialising in advanced user interfaces for information access and search, and R&D group manager at Canon Research Centre Europe, where he led a team developing next generation information access products and services. He currently holds an Honorary Visiting Fellowship at the Centre for Interactive Systems Research, City University London. He can be contacted via tgr@uxlabs.co.uk.

Book review:
A Generative Theory of Relevance

By Victor Lavrenko and W. Bruce Croft (Eds.)
Reviewed by Jianhan Zhu

Relevance is a fundamental concept in information retrieval. When we are searching for information, we are constantly judging a piece of information based on its relevance to our task or need at hand. Therefore, when search engines are filtering information on the users' behalf, they must have a good sense of a document's relevance to a user information need. On the Web, when the search queries are typically short, it is a real challenge for search engines to make sense of the queries in order to producing a ranked list of documents of potential relevance to the users.

Since the early years of information retrieval, researchers were constantly trying to model and define the concept of relevance. There have been many different approaching for defining relevance. Different definitions of relevance result in different information retrieval models. Therefore, for anyone new to information retrieval, relevance is an important concept to start with, so he/she can have a better understanding of the field.

This book provides an excellent literature review of the history of relevance, which should be a good beginning for starters. The book then goes on to define a rather novel approach of defining and modeling relevance. This approach unifies user information needs and documents within the same framework. These information needs and documents can be in different forms such as queries, different languages, handwritten information, images, and videos etc.

Chapter 1 introduces the new relevance model, and summarizes the contributions.

Chapter 2 gives a rather conclusive literature review of previous research on relevance such as user oriented and logical views of relevance, binary nature of relevance, and dependent and independent relevance. Based on the different definitions of relevance, there are different models such as probabilistic model and language models proposed for ranking documents.

Chapter 3 introduces the central idea of the book that both user information needs and documents are randomly generated from the same underlying relevance model. Although a short user query and a long document look very different from each other, based on the author's relevance model, they might be generated from the same underlying model. Therefore, the document is relevant to the query. The author casts the retrieval process as a hypothesis testing of whether a document and a query are generated from the same relevance model.

Chapter 4 devotes to different ways of obtaining the underlying relevance model. The author gave an overview of five existing generative models including pLSI (probabilistic latent semantic indexing) and LDA (latent Dirichlet allocation) etc. The author then showed that these models can all be reduced to the same simple form. Finally, the authored proposed a new generative model with some desirable properties.

Chapter 5 applies the relevance model to a number of retrieval scenarios including ad hoc retrieval, relevance feedback, cross language retrieval, handwriting retrieval, image retrieval, video retrieval, structured search with missing data, and topic detection and tracking. This shows that the proposed model provides a unified framework to a number of retrieval problems. The chapter also shows that the model works well in experiments on standard data sets.

Chapter 6 concludes the book and proposed future research including relevance based indexing, application to linked and relational data, semi-structured data, order-dependent data, and Dirichlet kernels.

Overall, this book can serve as a good companion for information retrieval students.

The book can help them gain a good understanding of the core concept, relevance, in information retrieval. The book will also help them study how to design effective and theoretical information retrieval models. The readers should have some background knowledge in statistics and probability theory in reading chapter 3 and 4.

Jianhan Zhu is a research fellow working on information retrieval, Web search, and text mining at the University College London, UK. Before that, he worked as a research fellow at the Knowledge Media Institute of the Open University, UK. He has a PhD in Computer Science from University of Ulster, UK, and a BSc in Automatic Control from Tsinghua University, China. He is a member of the ACM.

He can be contacted via: j.zhu@cs.ucl.ac.uk

Book Review:

Introduction to Video Search Engines

By David C Gibbon and Zhu Liu

Reviewed by Nilesh Thatte

The intended audience of this book are undergraduate students of computer science, developers and professionals working in the area of multimedia information retrieval. It covers the overall process of content acquisition, indexing and retrieval with examples of systems from academic and commercial world.

Authors (David C Gibbon and Zhu Liu) are experts in the field of multimedia content processing, content analysis, pattern recognition and machine learning jointly holding more than 20 US patents and around 40 pending patents. They have made significant contribution to MIRACLE (Multimedia Information Retrieval by Content) system, an AT & T platform for media processing algorithm development. MIRACLE system is discussed end to end as a case study in Chapter 10.

The book is conveniently organised in three sections covering fundamentals of video search engines, media processing and case studies respectively. Chapters follow a logical flow from content acquisition to display of results.

Chapter 1 steps straight into the heart of the multimedia web by presenting the challenges and opportunities of extracting information from video with technical advancement like ubiquitous broadband, IPTV and HD content. Today video content is not only easily accessible to masses but users are actively contributing to the pool by producing video content at low cost.

Chapter 2 discusses metadata standards and attributes of different types which could be leveraged by search engines. Presently there are various metadata and timed text formats in use as there is no global agreement on these standards. It was highlighted that availability of wide range of content sources and their own associated description formats is a challenge for video search engines. Though there was an emphasis on the US broadcasting media and standards.

Technical aspects of digital video are presented in Chapter 3. By nature of the content this chapter is more technical than previous ones. It discusses the challenges video search engines face in meeting the expectations of users of broadcast media and issues it must deal, like video aspect ratio, video formats, video compression, transport media, DRM and encoding.

High level architecture of video search engines is like a typical information retrieval system. It is explained in the concluding chapter (Chapter 4) of section one. Three main processes content acquisition, indexing and retrieval are followed by discussion of user perspective, scalability and retrieval interfaces. User perspective is covered as it poses challenges in designing the user interfaces.

Section 1 discusses Metadata as today most of the public interfacing video search engines are metadata retrieval systems. Section 2 covers various methods to extract meaningful information from the multimedia content with a possibility to build some structure and semantics about the content. This section covers the techniques that could possibly employed by search engines to improve the effectiveness.

Chapter 5 is an introduction to some media processing operations focusing on there usability for video search engines. Feature extraction is described as a stage to generate and select the features to reduce the data while retaining meaningful information. Media segmentation process could be used to divide the video stream in semantically consistent units benefiting in more accurate information retrieval. Clustering and data driven processing methods are discussed for high quality of automatic labelling.

Chapter 6-8 focuses specifically on processing technologies for visual, audio and text streams

of content. Techniques to analyse the visuals in video content are discussed in chapter 6. Shot boundary determination techniques developed in AT & T labs are described in detail. More media processing algorithms for Representative Image Selection, Face Detection, Face Recognition, Optical Character Recognition and Concept Detection using Images are discussed with mathematical formulas and relevant graphs. Considering presentation of video results is important for richer user experience few different ways to present the video content are also discussed focusing on the outputs from previously discussed processing methods.

Media processing techniques are discussed in context to audio stream in Chapter 7. Feature Extraction, Audio Segmentation, Speech Recognition and Content Categorization are discussed in detail to analyse the audio contents using Speech Signal Processing. Automatic speech recognition is a very useful processing technique for search engines. There are systems which work on acoustic models to convert a stream of audio into a sequence of words and have achieved more than 80% accuracy depending on conditions. This is an active area of research in many academic and industrial laboratories. Audio analysis is language dependent and in my opinion poses more challenges than visual processing.

Chapter 8 discusses text processing techniques which provide crucial cues for understanding contents. Many of the techniques discussed are mature and widely used in information retrieval systems. Part-of-Speech Tagging, Story Segmentation and Named Entity Extraction algorithms are specifically useful for Video Search Engines.

Chapter 9 discusses analysing different constituents of video together with example case studies. It covers Closed Caption Alignment, Multimodal Content Segmentation and Major Cast Detection. In the last chapter of this section authors describes the advantages of multimedia processing over isolated processing. MIRACLE system is used as an example of Closed Caption Alignment.

Last section of the book comprise of 2 chapters. Chapter 10 introduces various

academic and industrial research projects followed by early internet deployments and media monitoring systems. A list of multimedia retrieval dataset is provided for people interested in pursuing research in multimedia retrieval. An end to end system (MIRACLE System) is discussed in the case study to convey the computational requirements of a Video Search Engine.

Chapter 11, the concluding chapter covers current trends in the Video Search. Authors have rightly stayed away from predicting the future in rapidly growing technology though they have discussed various new video production, distribution and consumption methods. Content Segmentation, Summarisation, Time-based Metadata and Semantic Video Search are identified as ongoing research areas.

Glossary is provided at the end but many abbreviations are missing from the list. An exhaustive list would have been very helpful.

Overall the book meets its stated objective of introducing the reader to techniques behind Video Search Engine. It has covered various areas in the field of image processing, computer vision, speech recognition and information retrieval. It is well organised with lots of references for further reading. I would recommend this book for intended audience of students and developers though not much for the experts working in this area.

Nilesh Thatte is Senior Search Specialist at ioko. He holds Masters in Management Science, specializing in Information Systems. He is also pursuing a Masters in Intelligent Information Systems from Birkbeck College. His interests include IR, Video Search, Data mining and Clustering. Email:
nileshthatte@yahoo.com

Informer

Forthcoming Events

Edited By Andy MacFarlane

Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-2010)
A video processing conference of interest to members working on video search.
Scottsdale, Arizona, U.S.A, 13th to 15th January 2010.
http://www.vpqm.org/

Second IEEE International Conference on Intelligent Human Computer Interaction (IHCI 2010)
A general HCI conference of interest to members working on the interface size of search.
Allahabad, India, 16th to 18th January 2010.
http://hci.iiita.ac.in/

Dutch-Belgian Information Retrieval Workshop (DIR 2010)
Annual 'benelux' workshop on IR.
Nijmegen, Netherlands, 25th to 26th January 2010.
http://www.ru.nl/ds/ifl/dir_2010/

Third International Conference on Web Search and Data Mining (WSDM 2010)
A conference with a theme of interest to members working on web search.
New York City, U.S.A, 4th to 6th February 2010.
http://www.wsdm-conference.org/2010/

The Second International Conference on Information, Process, and Knowledge Management (eKNOW 2010)
A general conference on knowledge management of interested to members working in the area of KM and search.
St. Maarten, Netherlands Antilles, 10th to 15th February 2010.
http://www.iaria.org/conferences2010/eKNOW10.html

The Eleventh Workshop on Mobile Computing, Systems, and Applications (HotMobile 2010)
A general mobile computing conference, of interest to members working on mobile search.
Annapolis, MD, US, 22nd to 23rd February 2010
http://www.hotmobile.org/2010/

SpokenQuery Voice Search Workshop 2010
A workshop of interest to members working on different forms of query entry.
Dallas, TX, USA, 13th March 2010
http://www.spokenquery.org/

The 25th ACM Symposium on Applied Computing - Special Track on: INFORMATION ACCESS AND RETRIEVAL
A general applied computing conference, with a track on search.
Sierre, Switzerland, 21st to 26th March 2010.
http://www.disco.unimib.it/go/Home/Italiano/IAR-2010

Data Compression Conference (DCC 2010)
A conference focusing on compression of interest to members working in the area of compression and indexes.
Snowbird, Utah, USA, 24th to 26th March 2010
http://www.cs.brandeis.edu/~dcc/

32nd European Conference on Information Retrieval (ECIR 2010)
The annual conference of the IRSG.
The Open University, Milton Keynes, 28th to 31st March 2010.
http://kmi.open.ac.uk/events/ecir2010/

11th ACM International Conference on Multimedia Information
A general multimedia conference with a special session on music search.
Philadelphia PA USA, 29th to 31st March 2010
http://labrosa.ee.columbia.edu/3mir09/

International Academic Research Conference ACM-BCS Visions of Computer Science 2010
A general computer science conference of interest to members with a number of different interests in search.
Edinburgh, Scotland, 13th to 16th April 2010.
http://www.bcs.org/server.php?show=nav.11980

19th International World Wide Web Conference (WWW2010)
The conference on the web, of interest to members working in web search.
Raleigh, North Carolina USA, 26th to 30th April 2010.
http://www2010.org/www/

Informer

9th RIAO Conference Adaptivity, Personalization and Fusion of Heterogeneous Information (RIAO 2010)
Highly regard triennial conference, focusing on all aspects of information processing.
Paris, France, 28th to 30th April, 2010.
http://riao.free.fr/index.php?lang=en

Tenth SIAM International Conference on Data Mining (SDM2010)
Of interest to members working in the area of text mining.
Columbus, Ohio, USA, 29th April to 1st May 2010.
http://www.siam.org/meetings/sdm10/

BCS
INFORMATION
RETRIEVAL

Featured Job: **Senior Product Manager (UI, UE, UX, Search Front End)**

London / £50000 - £65000 per annum + bonus benefits

Working for a global retail organisation, the Senior Product Manager / User Interface Manager / User Experience Manager (UI, UE, UX, Web Front End) will develop business requirements and specify front end products to improve the customer experience and customer feedback, driving business growth.

The Senior Product Manager / User Interface Manager / User Experience Manager (UI, UE, UX, Web Front End) will understand consumer needs, behaviour and requirements across EU territories as well as the business and competitive context, reflecting this in business requirements for front end design initiatives; own the Search front end experience, Search Results Page designs, Navigation and User experience simplification; influence the global product lifecycle on all search front end initiatives to ensure global products meet EU requirements; directly product manage EU-specific projects and initiatives; work with the marketing organisation to define appropriate communications of product changes and launches to customers and work the analytics organisation to understand product performance and identify corrective actions.

Candidates for the Senior Product Manager / User Interface Manager / User Experience Manager (UI, UE, UX, Web Front End) role must have strong experience of online consumer Product Management coupled with strong understanding of appealing front end UI, GUI, UE design and UI design best practice. There are excellent opportunities to take on more responsibility and fast track your career.

Keywords: Senior Product Manager / User Interface Manager / User Experience Manager (UI, UE, UX, Web Front End)
Reference: 0910-09
Company Name: Intelligent People Ltd
Contact Name: Gemma Hall
Contact Telephone: 01727 736698

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.