

SIGIR 2005 Doctoral Consortium

Chair: David J Harper
Smart Web Technologies Centre
The Robert Gordon University
St Andrew Street, Aberdeen AB25 1HG, UK
djh@comp.rgu.ac.uk

INTRODUCTION

The second SIGIR Doctoral Consortium is being run as part of SIGIR 2005 in Salvador, Brazil. The purpose of the Consortium is to provide doctoral students with a forum, in which to discuss their PhD thesis proposal with experienced information retrieval (IR) researchers and other doctoral students. An important motivation in establishing the Consortium was to enrich the level of interaction between students and experienced IR researchers, and especially for those students from developing research groups. The Consortium also provides an opportunity for students to establish a supportive community of other doctoral students at the same stage as themselves.

Student participation is based on a submitted paper, which was reviewed by the Consortium Program Committee (CPC). In this second year of operation, we received 17 submissions, of which 10 were accepted. Student participants study in five countries, and present a broad range of perspectives on our field.

The process for submitting, selecting and publishing Consortium papers is as follows. Prospective student attendees submitted a four page technical paper, together with a one page “case for attending” prepared by the student and his/her advisor. Each submission was reviewed by two CPC members, based on technical considerations and on the perceived benefit of attendance. Abstracts of the accepted papers are included below as part of this paper. Students were also encouraged to submit posters or full papers based on the Consortium paper to the main SIGIR Conference, and some students took advantage of this opportunity for wider dissemination of their work.

The event itself takes place on the Tutorials Day of the SIGIR Conference. The format of the day is: short presentations of the papers with discussion in plenary session; detailed discussion between each student and two members of the CPC; and social interaction over lunch and dinner.

Financial support for student travel was generously provided by the ACM SIGIR. The SIGIR Conference Chair kindly agreed to cover the running costs of the Consortium. We also acknowledge Springer Science+Business Media for supporting the “Doctoral Consortium Award”.

The Doctoral Consortium is a permanent event within the annual SIGIR Conference, and is contributing to the development of new researchers in the field, and hence to the growth of the international IR community. The Consortium is evolving based on our experiences in running the event, and on feedback from the participants and the community.

Finally, as Chair, I would like to express my profound thanks to the members of the Consortium Program Committee for their efforts in reviewing both papers and abstracts, and for participating in the Consortium.

They are:

- Nick Belkin, *Rutgers University, USA*
- Jamie Callan, *Carnegie Mellon U., USA*
- Bruce Croft, *University of Massachusetts, USA*
- Sue Dumais, *Microsoft Research, Redmond, USA*
- Liz Liddy, *Syracuse University, USA*
- Yoëlle Maarek, *IBM Research, Israel*
- Alistair Moffat, *University of Melbourne, Australia*
- Doug Oard, *University of Maryland, USA*
- John Tait, *University of Sunderland, UK*
- Fabrizio Sebastiani, *University of Padua, Italy*

An Analysis of Evolved Term-weighting Schemes in Information Retrieval

Ronan Cummins
Dept. of Information Technology
National University of Ireland
Galway, Ireland
ronan.cummins@nuigalway.ie

Abstract

Machine Learning techniques are increasingly being applied to many areas in Information Retrieval. Evolutionary computation and Genetic Programming in particular have been shown to be a viable alternative to other standard analytical methods for developing term-weighting schemes in IR. This paper presents term-weighting schemes that have been evolved in both a global (collection-wide) and local (within-document) context.

In particular, global term-weighting schemes are evolved which have characteristics similar to that which Luhn predicted would lead to identifying terms with a high resolving power. Local (within-document) weighting schemes are evolved dependent on the best performing global scheme and we show an increase in mean average precision over the BM25 scheme for the full evolved scheme (i.e. the combined local and global scheme). An analysis of the term-frequency influence of best performing within-document scheme is shown to behave similarly to that of *Okapi-tf* when its term-frequency influence parameter is assigned a low value.

The document normalisation part of the evolved local scheme does not perform as well as *Okapi-tf* on long documents. We conclude that *Okapi-tf* can be tuned to interact effectively with the evolved global weighting scheme presented and increase average precision over the standard BM25 scheme.

Score Region Algebra: A Framework for Structured IR

Vojkan Mihajlovic
CTIT, University of Twente
P.O. Box 217, 7500AE Enschede, The Netherlands
vojkan@cs.utwente.nl

Abstract

We address the problem of developing a flexible framework for information retrieval (IR) in structured documents, such as XML. The framework is able to support a wide range of structured IR queries, transparent instantiations of different retrieval models, and different physical implementations. It is based on so-called *score region algebra (SRA)* that can express the following four essential ranked retrieval aspects for structured IR: term and element selection, element relevance score computation, element score propagation, and element score combination. Our preliminary research shows that different instantiations of each aspect, as well as different combinations of these instantiations, yield significantly different results. Our goal is to better understand structured IR by studying these aspects alone and their combination in the framework of SRA, and to use this knowledge to improve our structured IR system.

Advancing Corpus-Based Bilingual Retrieval

Paul McNamee
University of Maryland
Baltimore County, USA
mcnamee@cs.umbc.edu

Abstract

Bilingual dictionaries are widely available and are the most frequently used type of translation resource for bilingual retrieval. However, even among the most commonly spoken languages, for which large bilingual lexicons exist, dictionary-based translation suffers from several significant problems. These include: difficulty handling proper names, which are often missing; issues related to morphological variation, since entries or query terms may not be lemmatized; and, an inability to robustly handle multiword phrases, especially non-compositional expressions.

This paper outlines a plan of research to investigate how two techniques based on parallel texts can address these problems. To combat the problems of out-of-vocabulary words and morphological variation the use of *subword translation*, statistical translation of character n-grams rather than words or stemmed words, will be explored. To enable translation of non-compositional compounds and arbitrary length phrases, a recently developed method for efficiently determining frequencies of occurrence of such phrases in large corpora will be used. With these frequencies it is anticipated that statistical translation of complex phrases can be improved. Preliminary results are presented that suggest that both of these techniques can enhance bilingual retrieval performance.

The Re:Search Engine: Helping People Return to Information on the Web

Jaime Teevan
Massachusetts Institute of Technology
32 Vassar Street, G472, Cambridge, MA 02139 USA
teevan@mit.edu

Abstract

Re-finding information is commonly cited as a problem on the Web. One reason re-finding on the Web is difficult is that while people rely on a considerable amount of context to return to information (e.g., the original path taken to it), the Web makes no guarantee that the context will remain static. The Re:Search Engine is a system built to help people return to information in the dynamic environment of the Web by maintaining consistency in the search results it returns across time. For example, if Connie, while looking to purchase a Global Positioning System, found several systems she liked via a search for “GPS”, she would expect to be able to use the same query to locate the exact same systems again. However, simply returning the original result list when she re-issues the query might omit newly available GPS systems that she would like to see. The ideal result list would contain both the systems Connie remembers as having seen and high quality new systems. Because people tend not to remember much of what is presented in a search result list, when a person repeats a query, the Re:Search Engine is able to preserve what is remembered about the original result set while still presenting new information.

Web Information Access in the Blog Age: Opinions Matter

Gilad Mishne
Informatics Institute
University of Amsterdam
Kruislaan 403, 1098SJ Amsterdam
The Netherlands
gilad@science.uva.nl

Abstract

With the increase in the web's accessibility to the masses in the last years, the profile of web content is changing: more and more web pages are authored by non-professionals. Part of this "publishing revolution" is the phenomenon of blogs – personal, highly dynamic, and opinionated journals that are publicly available on the internet. While this domain has been fertile ground for research in the social sciences, it received relatively little attention from the retrieval community.

The subjective nature of blogs creates new challenges in retrieval and data mining. What are the user information needs expressed in this domain? How do we define the notion of relevance for addressing these needs? What knowledge can be mined from the structure and contents of blogs, and how?

In the proposed research, I intend to examine retrieval from blogs and additional domains rich in subjective contents (e.g., online forums). I will explore notions of non-topical relevance, constructing a corpus of annotated blogs for this purpose. I will characterize the types of information needs presented by users, and experiment with different ranking models for addressing these needs. The models will include linguistic and statistical features, but also non-content features such as recency. Finally, I will seek to uncover new knowledge which is hidden in the contents and the structure of subjective web documents, addressing tasks such as detecting trends and their propagation throughout the domain, classifying sentiments, identifying the community structure and so on. Both content and link analysis methods seem beneficial for this type of data mining as well as for retrieval.

Recovering Chronological Relationships in Dynamic Information

Jahna Otterbacher
School of Information
University of Michigan
1075 Beal Avenue, Ann Arbor, MI 48109-2122 USA
jahna@umich.edu

Abstract

An important feature of emergency news stories published online over time is their dynamic nature. As what is known about a situation changes, articles are updated to reflect new information. Currently, these dynamics are modeled using a phylogenetic technique, in which the evolutionary history of a set of texts is

reconstructed. The assumption is that stories evolve over time, beginning with the initial text published. "Mutations" occur as writers learn more about the story and update the previous version of the text. If an evolutionary model fits the data, it can then be used to infer chronological relationships between texts. The goal of the current experiments is to determine how to produce phylogenetic trees describing the evolution of reported answers to factual questions over time. A corpus of stories, such as plane crashes and natural disasters, was created. Annotators read all articles about a given story, and generated a list of related factual questions. For each story, multiple judges found all sentences providing an answer to each question. Given a set of sentences containing answers to a given question, pairwise distances between them will be found using an appropriate similarity metric. A phylogenetic method then uses these distances to construct a tree. If the model fits well, the sentences should be ordered in the tree chronologically. This hypothesis will be evaluated by comparing the algorithm's ordering of the sentences to the actual publication times of the source documents. Kendall's rank-order correlation coefficient will be used to test for statistical significance.

Focused Information Retrieval from Semi-Structured Documents

Börkur Sigurbjörnsson
Informatics Institute
University of Amsterdam
borkur@science.uva.nl

Abstract

When a user is confronted with a list of relevant documents, in response to a query, her search task is not over. She still has to explore the documents in the list in order to get to the relevant information. When documents are long, this can turn out be a tedious task for the user. It is thus desirable if the retrieval system can give the user a more focused access to the relevant documents by giving a more direct access to the relevant parts within the relevant documents. This task is twofold. First of all, the system needs to identify the relevant subparts of the documents. Second, the system needs to display these sub-document results to the user.

I will address both tasks in the context of retrieving information from semi-structured (XML) documents. XML documents have the advantage that they are divided into a presumably meaningful hierarchy of retrievable elements. However, the element retrieval is difficult to evaluate since the elements overlap each other. I will evaluate my element retrieval methods using the INEX and HARD test collections, where relevance is assessed at the sub-document level. Both collections have many open questions regarding the evaluation methodology. By looking at the two collections in parallel I hope provide better understanding of the nature of the search tasks being evaluated. For the result representation task I will evaluate how the element retrieval can be used to improve the result representation. The evaluation will be performed through user studies.

A Quality Focused Crawler for Health Information

Thanh Tin Tang
Department of Computer Science
Australian National University
Canberra, ACT 0200, Australia
thanh.tang@anu.edu.au

Abstract

Web users increasingly employ search engines to locate medical information. Unfortunately, the health information they obtain may be of poor quality or dubious value. We consider a specific search application that raises three interesting general issues. The application is a web search system geared towards locating high-quality medical advice, in the area of mental health. The three issues are: how to define and evaluate page quality; how to automatically detect quality; and how to build a domain-specific search engine integrating both quality and domain relevance measures.

Our initial experiment judged quality against the available scientific evidence: do search engines return results consistent with evidence-based medicine? Results showed that BluePages, a depression-specific index of 207 hand-picked sites without using expert knowledge on quality, gave slightly worse relevance results to Google but better quality advice. Analysis suggested that some of the relevant content returned by Google was linked to from the 207 sites and the better performance by Google was due to its better coverage, so it would make sense to expand the coverage of our depression-specific engine by following links.

Therefore the goal is to build a crawler that focuses on quality and domain relevance in order to find pages which are on topic (mental health related) while minimising bad advice. This will require extending published focused crawling methods to consider quality, and in general to study quality indicators on the Web.

Query-based, task-specific automatic summarization of biomedical literature on gene-gene relations

Jianji Yang
Department of Medical Informatics and Clinical
Epidemiology
Oregon Health & Science University
3181 S.W. Sam Jackson Park Road,
Portland, OR 97239-3098, USA
yangj@ohsu.edu

Abstract

With the increasing volume of published scientific articles, even the most robust IR engine returns more documents than biomedical scientists are able to handle. The problem is aggravated by the information-intensive nature of some “high-throughput” technologies, e.g. microarray experiments that can study gene expression at a genome-wide scale. In fact, these

advanced technologies and the increased number of publications discussing the findings impair our ability to fully comprehend the meanings that are embedded in the vast body of free texts in the literature. One possible way to alleviate the burden is to automatically produce customized summaries from the articles of interest for the users. Automatic summarization systems have been studied and applied in various domains with success. But adapting the technology in biomedicine domain has some major challenges, e.g., non-standard nomenclatures for genes, proteins and other biological entities, domain specific languages and the highly complex interrelation within biological systems. In this attempt to study how we can effectively summarize biomedical literature, I propose to build a query-based, task-specific automatic summarizer for information on gene-gene relations by applying machine learning and natural language technologies. The major contribution will be a summarizer that will incorporate a filter for different types of literature and will highlight contradictions in source articles. By initially restricting the problem domain and using real users to do evaluation, I hope to build an accurate summarizer that can be actually used by molecular biologists in everyday research.

Topic Hierarchy Construction for Web Site Summarization

Yongzheng Zhang
Faculty of Computer Science
Dalhousie University, Canada
6050 University Ave.
Halifax, Nova Scotia
Canada, B3H 1W5
yongzhen@cs.dal.ca

Abstract

Web document summarization, which identifies the essential contents of source documents, has gained much attention in recent years. In our previous work, we proposed a content-based system to generate a concise Web site summary by means of key phrase and key sentence extraction. Experimental evaluation shows that the automatically generated summaries can convey the same information as human-authored summaries do. However, as the size and diversity of Web sites continue to grow, straightforward summarization of an entire Web site may lead to an incoherent summary or a summary heavily biased to a subset of main topics covered in the target Web site. In this paper, we propose a new approach to summarization of Web sites with diverse topics and heterogeneous contents. The system is focused on creating a topic hierarchy which effectively organizes and visualizes the Web site contents. First a proper link hierarchy is constructed by capturing and refining the raw hyperlink structure of a given Web site. Second, documents in the link hierarchy are clustered into topic groups in a top-down manner. Third, topic groups are summarized and labeled to form a topic hierarchy in a bottom-up manner. We aim to apply extrinsic evaluation to measure the usefulness and effectiveness of the proposed approach in terms of how well the hierarchical summaries can help Web users understand the main topics and essential contents of Web sites.

Tutorials

Digital Libraries: An Overview and Formal Framework

Ed Fox & Marcos Gonçalves

Bioinformatics and Genomics for Information Retrieval

William Hersh & Hugh Williams

Multimedia Retrieval: Audio, Speech, Images, and Video

Dulce Ponceleon & Malcom Slaney

Introduction to Web IR

Andrei Broder & Prabhakar Raghavan

Introduction to Logistic Regression

David Lewis

The Continued Saga of DB-IR Integration

Ricardo Baeza-Yates & Mariano Consens

Statistical Language Models for Information Retrieval

ChengXiang Zhai

Workshops

ELECTRA: Methodologies and Evaluation of Lexical Cohesion

Organizers: Olga Vechtomova, Rosie Jones, Gael Dias

In this workshop we are interested in pointing at successes and failures of the integration of lexical cohesion in real-world IR applications such as document and passage retrieval, question answering, topic segmentation and text summarization. On the one hand, lexical cohesion has received much attention in Information Retrieval research during its more than 30-year old history, but so far with mixed results. On the other hand, a considerable amount of research has been devoted to this subject, both in terms of theory and practice, by the Natural Language Processing community, but with limited evaluation in real-world applications. This workshop is intended to bring together IR and NLP researchers and discuss what has been achieved in this area, to establish common themes between different approaches, and to discuss future research directions.

Heterogeneous and Distributed Information Retrieval

Organizers: Ranieri Baraglia, Domenico Laforenza, Fabrizio Silvestri

This workshop will focus on new methods and algorithms to efficiently and effectively access data distributed over large heterogeneous distributed systems. The workshop particularly encourages papers that address the creation and the search in distributed, dynamic information systems as well as papers presenting novel architectural solution for these systems. However, more broadly, papers are solicited on any topic related to information retrieval in distributed architectures.

IRiX: Information Retrieval in Context

Organizers: Peter Ingwersen, Kalervo Jarvelin, Nick Belkin

Context implies commonly interactive IR. It is assumed that contextual data can be used effectively to constrain retrieval of information thereby reducing the complexity of the retrieval process. The challenge is to understand and capture relevant context features. The IRiX workshop will focus on three major lines of action that explore the central features or evidence of context: 1) What are the elements of context, which are potentially significant to IR? 2) Which of these elements are, or could be useful in improving IR? 3) How can features of context be used to improve IR? IRiX provides: Oral presentations on actions, selected background position papers and discussion group activities.

Multimedia Information Retrieval

Organizers: Alex Hauptmann, R. Manmatha, Stefan Rueger

This full-day workshop will include discussion and papers. The workshop solicits submissions on searching and retrieving images, speech, video, text and music or any combination of these; topics include but are not limited to: Content-based indexing, search and retrieval; Feature extraction and representation; Automated semantic annotation; Relevance feedback; Query models, paradigms and languages; Search and browsing mechanisms; Document classification; Evaluation; User studies; Meta-data (e.g., MPEG-7); Ontologies and Taxonomies; and Applications.

Mathematical/Formal Methods in Information Retrieval

Organizers: Sandor Dominich and Iadh Ounis

This workshop aims at promoting discussion and interaction among those with theoretical and applicative research interests in mathematical/formal aspects of Information Retrieval coming from a large spectrum of different IR fields, and also at being a forum for the presentation of both theoretical and applicative results (e.g., foundational issues; description and/or integration of models; retrieval applications; mathematical/formal techniques, properties and structures in IR; existing and/or new theories and theoretical aspects, interdisciplinary approaches).

Predicting Query Difficulty — Methods and Applications

Organizers: David Carmel, Ian Soboroff

Estimation of query difficulty is an attempt to quantify the quality of results returned by the search system for a query. Ideally, a system that can predict difficult queries can adapt parameters or change algorithms to suit the query. In this workshop we would like to explore techniques for prediction of and adaptation to query difficulty. We plan to focus on the reasons that cause a specific query to become difficult, classification of queries and failure modes, evaluation methodology, as well as on potential applications for query prediction.

Stylistic Analysis of Text for Information Access

Organizers: Shlomo Argamon, Jussi Karlgren, Jimi Shanahan

HOW something is expressed, as opposed to WHAT is expressed is a many-faceted and elusive, yet intuitively important and patent characteristic of human linguistic expression. This workshop will discuss issues in the automatic analysis and extraction of stylistic variation of natural language texts — especially but not exclusively addressing concerns related to information access. Major questions include: Style in Theory: What is style? Style in Engineering: How is style analyzable? Style in Applications: What tasks can stylistic information be used for? Style in Research: What tools and resources do you use, and can we use them too?

Author Index

- Agichtein, E. 567
Allan, J. 433
Anh, V. N. 226
Artiles, J. 569
Aslam, J. A. 27, 571, 573
Azzopardi, L. 575
- Badue, C. 577
Bai, J. 298
Bainbridge, W. 579
Barbosa, R. 577
Barreiro, A. 587
Bast, H. 11
Beitzel, S. M. 581, 583, 615
Bender, M. 67
Bennett, P. N. 585
Bernstein, Y. 678
Blanco, R. 587
Bodoff, D. 589
Bowden, M. 552
Boydell, O. 591
Brill, E. 567
Brisaboa, N. R. 234
Buntine, W. 647
- Cai, D. 3
Calado, P. 667
Callan, J. 83
Cano, P. 673
Cao, G. 298
Cao, Y. 250
Carbonell, J. 585
Carmel, D. 512
Carneiro, G. 559
Carterette, B. 433
Carvalho, V. R. 345
Chai, J. Y. 51, 536, 619
Chai, K. M. A. 593
Chakrabarti, S. 674
Chen, H.-H. 627
Chen, X. 266
Chen, Y. 667
Chen, Z. 114, 130, 186, 194
408, 504, 122
Cheng, Q. 122
Cheng, X. 653
- Chirita, P.-A. 178
Chowdhury, A. 581, 583, 615
Chua, T.-S. 384, 400, 651
Clarke, C. L. A. 314, 441
Cohen, W. W. 345
Craswell, N. 416
Cristo, M. 496, 667
Croft, W. B. 219, 472, 617, 678
Crowe, M. 575
Cucerzan, S. 567
Cui, H. 384, 400, 651
- Darlow, A. 512
Daumke, P. 675
Dave, S. 657
Davis, A. 657
de Moura, E. S. 496
de Vries, A. P. 282
Devlin, S. 629
Dollu, V. D. 457
Domshlak, C. 19
Dorr, B. J. 520
Downie, J. S. 676
Dumais, S. T. 449
- Efthimiadis, E. N. 595
Ehmann, A. F. 676
Elsayed, T. 597
Elson, D. K. 210
- Fan, W. 122, 130, 504, 667
Fang, H. 480
Fariña, A. 234
Ferguson, P. 599
Fetterly, D. 170
Fine, S. 512
Flores, F. 218
Forman, J. 424
Fox, E. A. 130, 667
Freund, L. 441
Frieder, O. 583, 615
Fu, X. 457
- Gao, J. 290
Gaussier, E. 601
Gay, G. 154
- Ghanem, M. M. 653
Ghoshal, A. 544
Girolami, M. 575
Golgher, P. 496, 577
Gonçalves, M. A. 667
Gong, Y. 274
Gonzalo, J. 569, 603
Goutte, C. 601
Granka, L. 154
Grivolla, J. 605
Grossman, D. 615
Guan, J. 671
Guo, Y. 653
Gurrin, C. 591, 599
- Hahn, U. 528
Hajič, J. 645
Harabagiu, S. 202
Harper, D. J. 639, 684
Hauptmann, A. 637
Hawking, D. 75
He, B. 465
He, X. 3
Hembrooke, H. 154
Hendy, D. G. 595
Hernandez, N. 607
Hirschberg, J. 210
Hoenkamp, E. 609, 611
Horvitz, E. 449
Hosanagar, K. 91
Howe, N. R. 377
Hu, G. 250
Hu, Y. 250
Huang, F. 669
Huang, X. 613
Huang, Y. R. 613
- Idreos, S. 322
Ircing, P. 544
- Jaakkola, T. 353
Jansen, B. J. 649
Jensen, E. C. 581, 583, 615
Jeon, J. 617
Ji, L. 504
Ji, X. 274

Jin, R.	51, 536, 619	Liu, T.-Y.	408	Paramá, J. R.	234
Joachims, T.	154	Liu, Y.	504, 536	Pass, G.	583
Jose, J. M.	35	Lu, Q.	631	Passonneau, R. J.	210
Kadobayashi, R.	621	Lu, Y.	194, 424	Pavlu, V.	27, 571, 573
Kajiyama, T.	623	Luk, R. W. P.	663	Perkiö, J.	647, 683
Kan, M.-Y.	384, 400, 651	Ma, B.	369	Peters, C.	603
Kando, N.	623	Ma, W.-Y.	106, 122, 408, 424 488, 504	Portinari, J. C.	1
Kankanhalli, M. S.	361	Maddage, N. C.	361	Poulain, S.	607
Kareem, O.	679	Magalhães, J.	641	Pouwelse, J.	659
Kaski, S.	146	Majumdar, D.	11	Puolamäki, K.	146
Kelly, D.	457	Manasse, M.	170	Qi, H.	290
Khudanpur, S.	544	Manmatha, R.	377	Qin, T.	408
Kisiel, B.	98	Markó, K.	528, 675	Radev, D.	679
Kohlschütter, C.	178	Masuyama, T.	338	Rath, T. M.	377
Kolcz, A.	581	Mayfield, J.	643	Reinders, M. J. T.	659
Koppenberger, M.	673	McKeown, K.	210	Rennie, J. D. M.	353
Koubarakis, M.	322	McNamee, P.	643	Ribeiro-Neto, B.	496, 577
Kraaij, W.	655	Medelyan, O.	528	Robertson, S.	416
Krishnan, V.	625	Metzler, D.	472, 678	Roelleke, T.	282
Ku, L.-W.	627	Michel, S.	67	Rüger, S.	641
Kurland, O.	19, 306	Mirchandani, J.	674	Ruthven, I.	35
Kwok, K. L.	663	Moffat, A.	226, 678	Salojärvi, J.	146
Lacatusu, F.	202	Moldovan, D.	552	Sanderson, M.	162
Lagendijk, R. L.	659	Monz, C.	520	Satoh, S.	623
Lee, J. H.	617	Mostafa, J.	138	Savia, E.	146
Lee, L.	19, 306	Mothe, J.	607	Schenkel, R.	242
Lee, L.-Y.	627	Murray, G. C.	635	Schulz, S.	528, 675
Lee, W. S.	266	Najork, M.	170	Sehgal, A. K.	680
Leidner, J. L.	677	Nakagawa, H.	338	Seki, K.	138
Lewis, D. D.	581	Nandi, A.	674	Shakery, A.	661
Lewis, J.	433	Narasimhan, M.	330	Shao, X.	361
Li, B.	631	Navarro, G.	234	Sharma, H.	649
Li, F.	633	Nejdl, W.	178	Shen, D.	194
Li, H.	250, 369, 504	Nenkova, A.	210	Shen, X.	43, 59, 681
Li, K.	400, 651	Neto, B. R.	496, 577	Sherfese, D.	682
Li, M.	106	Nie, J.-Y.	290, 298	Shi, S.	250, 488
Li, P.	589	Oard, D. W.	579, 597, 645, 665	Si, L.	83
Li, W.	631, 663	O'Driscoll, N.	682	Sibley, T.	657
Li, Y.	424	Olsson, J. S.	645	Simola, J.	146
Li, Z.	106	Otterbacher, J.	679	Singhal, A.	432
Liang, S. F.	629	Ounis, I.	465	Smeaton, A. F.	591, 599
Lin, C.	114	Paiu, R.	178	Smyth, B.	591
Lin, J.	392, 635	Pan, B.	154	Song, R.	250, 488
Lin, W.-H.	637			Srikanth, M.	552
Liu, B.	639			Srinivasan, P.	680
Liu, N.	122				

Strohman, T.	219	Verdejo, F.	569	Xue, G.-R.	114, 186
Sun, J.-T.	194	Viola, P.	330	Yan, J.	122, 130
Sun, R.	400, 651	Vogel, S.	669	Yan, S.	122
Tait, J.	629	Wack, N.	673	Yang, Q.	114, 186, 194
Tan, B.	43, 681	Wang, B.	106, 653	Yang, Y.	98, 633
Tan, S.	653	Wang, C.	424	Yilmaz, E.	27, 571, 573
Tanaka, K.	621	Wang, J.	659	Yom-Tov, E.	512
Tao, T.	661	Wang, L.	424	Yoo, S.	98
Taylor, M.	416	Wang, X.	661	Yu, K.	258
Tcheng, D.	676	Watt, S.	639	Yu, Q.	488
Teevan, J.	449	Weikum, G.	67, 242	Yu, S.	258
Theobald, M.	242	Wen, J.-R.	488	Yu, Y.	114, 186
Thomas, P.	75	Wen, M.	613	Zaragoza, H.	416
Tirri, H.	647, 683	White, R. W.	35, 579	Zeng, H.-J.	114, 186, 194
Toms, E. G.	441	Wilkins, P.	599	Zhai, CX.	43, 59, 480, 681
Tresp, V.	258	Wong, K. F.	663	Zhang, B.	122, 130, 504, 667
Triantafillou, P.	67	Wu, H. C.	663	Zhang, C.	51
Trieschnigg, D.	655	Wu, M.	631	Zhang, D.	266
Tryfonopoulos, C.	322	Wu, T.-H.	627	Zhang, J.	98
Turtle, H.	219	Wu, Y.	665	Zhang, X.-D.	408
Tuulos, V. H.	683	Xi, W.	114, 130, 504	Zhang, Y.	669
Udani, G.	657	Xia, X.	290	Zhou, S.	671
van Dijk, S.	609	Xie, X.	424	Zhu, S.	274
van Dinther, G.	611	Xin, G.	250	Zhuang, D.	130
Varner, J.	552	Xu, C.	361	Zimmer, C.	67
Vasconcelos, N.	559	Xu, H.	653	Ziviani, N.	577
		Xu, W.	274	Zobel, J.	162, 678