

Invited Talks

Time-Biased Gain

Charles L. A. Clarke (University of Waterloo)

Time-biased gain provides a unifying framework for information retrieval evaluation, generalizing many traditional effectiveness measures while accommodating aspects of user behaviour not captured by these measures. By using time as a basis for calibration against actual user data, time-biased gain can reflect aspects of the search process that directly impact user experience, including document length, near-duplicate documents, and summaries. Unlike traditional measures, which must be arbitrarily normalized for averaging purposes, time-biased gain is reported in meaningful units, such as the total number of relevant documents seen by the user. In work reported at SIGIR 2012, we proposed and validated a closed-form equation for estimating time-biased gain, explored its properties, and compared it to standard approaches. In work reported at CIKM 2012, we used stochastic simulation to numerically approximate time-biased gain, an approach that provides greater flexibility, allowing us to accommodate different types of user behaviour and increases the realism of the effectiveness measure. In work reported at HCIR 2012, we extended our stochastic simulation to model the variation between users. In this talk, I will provide an overview of time-biased gain, and outline our ongoing and future work, including extensions to evaluate query suggestion, diversity, and whole-page relevance. This is joint work with Mark Smucker.

The Challenges of Building Online Community Museums

Nigel Stanger (University of Otago)

Over the past few years I have been involved in the development of online museums for two different communities in Central Otago, New Zealand. Developing digital archives for the general public raises some interesting issues, including usability, copyright, curation of items, and potentially dealing with unusual document types. In this talk I will discuss these and other issues, and chronicle the development of both museums.

Symposium Chair

Andrew Trotman, University of Otago

Programme Co-chairs

Sally Jo Cunningham, University of Waikato
Laurianne Sitbon, Queensland University of Technology

Program Committee

Robert Allen, Victoria University of Wellington
Peter Bruza, Queensland University of Technology
Shane Culpepper, RMIT University
Sally Jo Cunningham, University of Waikato
David Eysers, University of Otago
Shlomo Geva, Queensland University of Technology
David Hawking, Funnelback
Yun Sing Koh, University of Auckland
Irena Koprinska, University of Sydney
Alexander Krumpholz, CSIRO/ ANU
Alistair Moffat, University of Melbourne
Glen Newton, Carleton University
Laz Park, University of Western Sydney
Luiz Augusto Pizzato, University of Sydney
Gitesh Raikundalia, Victoria University
Mark Sanderson, RMIT University
Falk Scholer, RMIT University
Laurianne Sitbon, Queensland University of Technology
Ling-Xiang Tang, Queensland University of Technology
James A. Thom, RMIT University
Andrew Trotman, University of Otago
Andrew Turpin, University of Melbourne
Mingfang Wu, RMIT University
Burkhard Wuensche, University of Auckland
Guido Zuccon, CSIRO

ADCS Steering Committee

Sally Jo Cunningham, University of Waikato
Shlomo Geva, Queensland University of Technology
Mark Sanderson, RMIT University
Falk Scholer, RMIT University
James A. Thom, RMIT University
Paul Thomas, CSIRO
Andrew Trotman, University of Otago
Andrew Turpin, University of Melbourne

Proceedings of the Seventeenth Australasian Document Computing Symposium
University of Otago, Dunedin, New Zealand
5-6 December 2012

Chair's Preface

These proceedings contain the papers of the Seventeenth Australasian Document Computing Symposium hosted by the University of Otago and held in Dunedin, New Zealand.

The quality of submissions was again very high this year. Of the 24 papers submitted, 11 were accepted for full presentation at the symposium (46%) and 8 were accepted for short presentation (32%). The full written version of each submission received three anonymous reviews by independent, qualified international experts in the area. Dual submissions were explicitly prohibited.

We would like to thank the members of the program committee for their reviewing efforts. We would also like to thank ACM SIGIR, Google, NICTA, Bing, Funnelback, and the University of Otago for their generous support of the event.

The symposium includes many formal presentations, but perhaps its greatest benefit lies in the opportunity it provides for document computing practitioners and researchers to get together and informally share ideas. Once again we have collocated with The Australasian Language Technology Workshop (ALTA), sharing a joint paper session, a poster session, and social events.

Andrew Trotman (Chair)
7 November 2012

Table of Contents

Full Papers

Effects of Spam Removal on Search Engine Efficiency and Effectiveness <i>Matt Crane, Andrew Trotman</i>	1..8
Efficient Indexing Algorithms for Approximate Pattern Matching in Text <i>Matthias Petri, J. Shane Culpepper</i>	9..16
Reordering an Index to Speed Query Processing Without Loss of Effectiveness <i>David Hawking, Timothy Jones</i>	17..24
Comparing Scanning Behaviour in Web Search on Small and Large Screens <i>Jaewon Kim, Paul Thomas, Ramesh Sankaranarayanan, Tom Gedeon</i>	25..30
Explaining Difficulty Navigating a Website Using Page View Data <i>Paul Thomas</i>	31..38
Relationship Between the Nature of the Search Task Types and Query Reformulation Behaviour <i>Khamsum Kinley, Dian Tjondronegoro, Helen Partridge, Sylvia Edwards</i>	39..46
Models and Metrics: IR Evaluation as a User Process <i>Alistair Moffat, Falk Scholer, Paul Thomas</i>	47..54
Sentence Length Bias in TREC Novelty Track Judgements <i>Lorena Leal Bando, Falk Scholer, Andrew Turpin</i>	55..61
Multi-Aspect Group Formation using Facility Location Analysis <i>Mahmood Neshati, Hamid Beigy, Djoerd Hiemstra</i>	62..71
An Ontology Derived from Heterogeneous Sustainability Indicator Set Documents <i>Lida Ghahremanloo, James Thom, Liam Magee</i>	72..79
Graph-Based Concept Weighting for Medical Information Retrieval <i>Bevan Koopman, Guido Zuccon, Peter Bruza, Laurianne Sitbon, Michael Lawley</i>	80..87

Short Papers

A Study in Language Identification <i>Rachel Milne, Richard O'Keefe, Andrew Trotman</i>	88..95
An Attempt to Measure the Quality of Questions in Question Time of the Australian Federal Parliament <i>Andrew Turpin</i>	96..103
An English-Translated Parallel Corpus for the CJK Wikipedia Collections <i>Ling-Xiang Tang, Shlomo Geva, Andrew Trotman</i>	104..110
Exploiting Medical Hierarchies for Concept-based Information Retrieval <i>Guido Zuccon, Bevan Koopman, Anthony Nguyen, Deanne Vickers, Luke Butt</i>	111..114
Finding Additional Semantic Entity information for Search Engines <i>Jun Hou, Richi Nayak, Jinglan Zhang</i>	115..122
Is the Unigram Relevance Model Term Independent? Classifying Term Dependencies in Query Expansion <i>Mike Symonds, Peter Bruza, Guido Zuccon, Laurianne Sitbon, Ian Turner</i>	123..127
Pairwise Similarity of TopSig Document Signatures <i>Chris De Vries, Shlomo Geva</i>	128..134
Putting the Public into Public Health Information Dissemination: Social Media and Health-related Web Pages <i>Robert Steele, Dan Dumbrell</i>	135..138