

ADCS 2008

Proceedings of the Thirteenth Australasian  
Document Computing Symposium

8 December 2008

Edited by

Rob McArthur, Paul Thomas, Andrew Turpin and Mingfang Wu



# **Proceedings of the Thirteenth Australasian Document Computing Symposium**

Tasmanian ICT Centre, CSIRO, Hobart, Tasmania  
8 December 2008

Published by  
School of Computer Science and Information Technology,  
RMIT University,  
Melbourne VIC 3001, Australia.

## Editors

Rob McArthur  
Paul Thomas  
Andrew Turpin  
Mingfang Wu

ISBN: 13978 1 921426 21 6

<http://es.csiro.au/adcs2008>

**Proceedings of the Thirteenth Australasian Document Computing Symposium**  
Hobart  
8 December 2008

## **Chair's Preface**

These proceedings contain the papers of the Thirteenth Australasian Document Computing Symposium hosted by the Tasmanian ICT Centre, CSIRO in Hobart, Tasmania.

The two keynote talks, seven long papers and three short papers reflect the breadth of interest of the Australian research community in the area of document computing. Continuing the tradition begun in 2007, ADCS is not only collocated with The Australasian Language Technology Workshop 2008, but we are sharing a paper session, keynote talk, and social functions with the Australian natural language research community.

The quality of submissions was very high this year. Of the 12 full papers submitted, 8 were accepted for presentation as full papers (67%), 1 was accepted as a short paper (8%), and 3 were rejected (25%). All of the 3 short papers submitted were accepted (100%).

All submissions received at least two anonymous reviews by experts in the area, while 60% of the papers received 3 reviews. Dual submissions were explicitly prohibited.

The members of the program committee and extra reviewers deserve special thanks for their professional reviews all received in the short time required for this ADCS conference.

We would also like to thank Funnelback, CSIRO, NICTA (Victoria) and RMIT School of Computer Science and IT for their generous sponsorship 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 and enthusiasm.

## **Symposium Co-Chairs**

Rob McArthur	CSIRO ICT
Mingfang Wu	RMIT University

## **Program co-chairs**

Paul Thomas	CSIRO ICT
Andrew Turpin	RMIT University

## **Program Committee**

Peter Bailey	Microsoft Research
Peter Bruza	Queensland University of Technology
Bob Colomb	University of Technology Malaysia
Stijn Dekeyser	University of Southern Queensland
Shlomo Geva	Queensland University of Technology
David Hawking	Funnelback Pty Ltd
Judy Kay	University of Sydney
Robert McArthur	CSIRO ICT
Alistair Moffat	University of Melbourne
Gitesh K. Raikundalia	Victoria University
Falk Scholer	RMIT University
Milad Shokouhi	Microsoft Research, Cambridge
Amanda Spink	Queensland University of Technology
Jamie Thom	RMIT University
Andrew Trotman	University of Otago
Anh Vo	University of Melbourne
Ross Wilkinson	CSIRO
William Webber	University of Melbourne
Justin Zobel	NICTA

## **Additional Reviewer**

Jing He	Victoria University
---------	---------------------

## **ADCS Advisory Committee**

Peter Bruza	Queensland University of Technology
Judy Kay	The University of Sydney
David Hawking	Funnelback
Alistair Moffat	The University of Melbourne
Amanda Spink	Queensland University of Technology
Ross Wilkinson	CSIRO, Canberra
Justin Zobel	NICTA

# Contents

<i>Chair's preface</i> .....	iii
------------------------------	-----

## Keynotes

<i>Quo vadis information retrieval research</i> .....	1
Kal Järvelin	
<i>Syntactic and semantic structure in web search queries</i> .....	2
Rosie Jones	

## Session 1

<i>Term-frequency surrogates in text similarity computations</i> .....	3
Stefan Pohl and Alistair Moffat	
<i>MetaView: Dynamic metadata based views of user files</i> .....	11
James Bunton, Judy Kay, and Bob Kummerfeld	
<i>On the relevance of documents for semantic representation</i> .....	19
Laurianne Sitbon and Peter Bruza	
<i>Exploring the benefit of contextual information for boosting TREC Genomic IR performance</i> .....	23
Bader Aljaber, Nicola Stokes, James Bailey, and Yi Li	
<i>WebKnox: Web knowledge extraction</i> .....	27
David Urbansky and Jamie Thom	

## Session 2

<i>Anonymous folksonomies for small enterprise webs: a case study</i> .....	35
Tom Rowlands, David Hawking, and Ramesh Sankaranarayana	
<i>The effect of using pitch and duration for symbolic music retrieval</i> .....	41
Iman S. H. Suyoto and Alexandra L. Uitdenbogerd	
<i>Extraction of named entities from tables in gene mutation literature</i> .....	49
Wern Wong, David Martinez, and Lawrence Cavedon	

## Session 3

<i>Facilitating biomedical systematic reviews using ranked text retrieval and classification</i> .....	53
David Martinez, Sarvnaz Karimi, Lawrence Cavedon, and Timothy Baldwin	
<i>Parameter sensitivity in rank-biased precision</i> .....	61
Yuye Zhang, Laurence A. F. Park, and Alistair Moffat	

## Session 4 (joint with ALTW)

<i>Querying linguistic annotations</i> .....	69
Sumukh Ghodke and Steven Bird	
<i>Using collaboratively constructed document collections to simulate real world object comparisons</i> .....	73
Karl Grieser, Timothy Baldwin, Fabian Bohnert, and Liz Sonenberg	

