

WILLIAM WEBBER

College of Information Studies, University of Maryland, United States
wew@umd.edu m: +1 301 906 3177

ACADEMIC RECORD

Ph. D. University of Melbourne, Australia, 2010. Thesis title: “Measurement in Information Retrieval Evaluation”. Supervisors: Alistair Moffat and Justin Zobel.

M. Eng. Sc. (Comp. Sc. & Soft. Eng.) (Research) University of Melbourne, Australia, 2006. Thesis title: “Design and Implementation of a Pipelined Distributed Information Retrieval Architecture”. Supervisors: Alistair Moffat and Justin Zobel.

Grad. Dip. Comp. Sc. RMIT University, Melbourne, Australia, 1997.

B. Arts (Honours) (First Class) University of Melbourne, Australia, 1992. Thesis title: “ ‘A Too Nice Inquisition after Truth’: William Chillingworth and the Church of England”. Supervisor: Donald Kennedy.

PROFESSIONAL EXPERIENCE

Research Associate August 2011 to July 2013
College of Information Studies, University of Maryland, United States
Research into the evaluation of e-discovery systems and processes, as part of NSF project #1065250, “Development and Evaluation of Search Technology for Discovery of Evidence in Civil Litigation”

Research Associate October 2006 to July 2011
Department of Computer Science and Software Engineering, University of Melbourne, Australia
Research into the evaluation of the performance of Information Retrieval systems.

Research Assistant February 2005 to September 2006
Department of Computer Science and Software Engineering, University of Melbourne, Australia
Research into distributed and parallelised query evaluation.

Research Assistant January 2004 to February 2005
School of Computer Science and Information Technology, RMIT University, Melbourne, Australia
Research into distributed and parallelised query evaluation.

Freelance Software Developer July 2001 to January 2004
Various
Web-centered development projects in Java and PHP.

Senior Programmer January 1999 to April 2001
PeopleWeb Communications, Melbourne, Australia
Development of multi-functioned web portal software for Silicon Valley-headquartered startup. Development in Java, using Servlets API.

Tutor, Senior Tutor March to December 1998
RMIT University, Melbourne, Australia
Tutor and senior tutor for postgraduate Algorithms and Data Structures and C Programming subjects.

Research Programmer July to December 1997
Multimedia Database Systems, RMIT University, Melbourne, Australia
C programmer on MG, a research full-text database.

Programmer

Cincom Systems of Australia, Melbourne, Australia
COBOL, 4GL, REXX programmer.

July to December 1996

PUBLICATIONS**Peer-reviewed conference, journal, and workshop**

Douglas W. Oard and William Webber. Information retrieval for e-discovery. *Foundations and Trends in Information Retrieval*, 2013. to appear.

William Webber and Jeremy Pickens. Assessor disagreement and text classifier accuracy. In Diane Kelly, Maarten de Rijke, and Tetsuya Sakai, editors, *Proc. 36th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland, July 2013. To appear.

Praveen Chandar, William Webber, and Ben Carterette. Document features predicting assessor disagreement. In Diane Kelly, Maarten de Rijke, and Tetsuya Sakai, editors, *Proc. 36th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland, July 2013. To appear.

William Webber, Mossaab Bagdouri, David D. Lewis, and Douglas W. Oard. Sequential testing in classifier evaluation yields biased estimates of effectiveness. In Diane Kelly, Maarten de Rijke, and Tetsuya Sakai, editors, *Proc. 36th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland, July 2013. To appear.

Falk Scholer, Diane Kelly, Hanseul S. Lee, Wan-ching Wu, and William Webber. The effect of threshold priming and need for cognition on relevance assessment. In Diane Kelly, Maarten de Rijke, and Tetsuya Sakai, editors, *Proc. 36th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland, July 2013. To appear.

William Webber. Approximate recall confidence intervals. *ACM Transactions on Information Systems*, 31(1): 2:1–33, 2013. doi: <http://dx.doi.org/10.1145/2414782.2414784>. URL <http://arxiv.org/abs/1202.2880v4>.

William Webber, Praveen Chandar, and Ben Carterette. Alternative assessor disagreement and retrieval depth. In *Proc. 21st ACM International Conference on Information and Knowledge Management*, pages 125–134, Maui, Hawaii, USA, October 2012a.

William Webber, Bryan Toth, and Marjorie Desamito. Effect of written instructions on assessor agreement. In William Hersh, Jamie Callan, Yoelle Maarek, and Mark Sanderson, editors, *Proc. 35th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1053–1054, Portland, Oregon, USA, August 2012b.

Justin Zobel, William Webber, Mark Sanderson, and Alistair Moffat. Principles for robust evaluation infrastructure. In *Proc. CIKM Data Infrastructures for Supporting Information Retrieval Evaluation*, pages 6:1–4, Glasgow, UK, October 2011.

William Webber. Re-examining the effectiveness of manual review. In *Proc. SIGIR Information Retrieval for E-Discovery Workshop*, pages 2:1–8, Beijing, China, July 2011.

William Webber, Douglas W. Oard, Falk Scholer, and Bruce Hedin. Assessor error in stratified evaluation. In *Proc. 19th ACM International Conference on Information and Knowledge Management*, pages 539–548, Toronto, Canada, October 2010a.

William Webber, Alistair Moffat, and Justin Zobel. The effect of pooling and evaluation depth on metric stability. In Tetsuya Sakai, Mark Sanderson, and William Webber, editors, *Proc. 3rd International Workshop on Evaluating Information Access*, pages 7–15, Tokyo, Japan, June 2010b.

William Webber, Alistair Moffat, and Justin Zobel. The effect of pooling and evaluation depth on metric stability. In Tetsuya Sakai, Mark Sanderson, and William Webber, editors, *Proc. 3rd International Workshop on Evaluating Information Access*, pages 7–15, Tokyo, Japan, June 2010b.

William Webber, Alistair Moffat, and Justin Zobel. A similarity measure for indefinite rankings. *ACM Transactions on Information Systems*, 28(4):20:1–38, November 2010c.

Timothy G. Armstrong, Alistair Moffat, William Webber, and Justin Zobel. Improvements that don't add up: Ad-hoc retrieval results since 1998. In David Cheung, Il-Yeol Song, Wesley Chu, Xiaohua Hu, Jimmy Lin, Jiexun Li, and Zhiyong Peng, editors, *Proc. 18th ACM International Conference on Information and Knowledge Management*, pages 601–610, Hong Kong, China, November 2009a.

William Webber and Laurence A. F. Park. Score adjustment for correction of pooling bias. In James Allan, Javed Aslam, Mark Sanderson, ChengXiang Zhai, and Justin Zobel, editors, *Proc. 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 444–451, Boston, Massachusetts, USA, July 2009.

Timothy G. Armstrong, Alistair Moffat, William Webber, and Justin Zobel. Has adhoc retrieval improved since 1994? In James Allan, Javed Aslam, Mark Sanderson, ChengXiang Zhai, and Justin Zobel, editors, *Proc. 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 692–693, Boston, Massachusetts, USA, July 2009b.

Timothy G. Armstrong, Alistair Moffat, William Webber, and Justin Zobel. EvaluatIR: An online tool for evaluating and comparing IR systems. In James Allan, Javed Aslam, Mark Sanderson, ChengXiang Zhai, and Justin Zobel, editors, *Proc. 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 834, Boston, Massachusetts, USA, July 2009c.

Timothy G. Armstrong, Justin Zobel, William Webber, and Alistair Moffat. Relative significance is insufficient: baselines matters too. In Shlomo Geva, Jaap Kamps, Carol Peters, Tetsuya Sakai, Andrew Trotman, and Ellen Voorhees, editors, *Proc. SIGIR 2009 Workshop on the Future of IR Evaluation*, pages 25–26, Boston, Massachusetts, USA, July 2009d.

William Webber, Alistair Moffat, and Justin Zobel. Statistical power in retrieval experimentation. In James G. Shanahan, Sihem Amer-Yahia, Yi Zhang, Alek Kolcz, Abdur Chowdhury, and Diane Kelly, editors, *Proc. 17th ACM International Conference on Information and Knowledge Management*, pages 571–580, Napa, California, USA, October 2008a.

William Webber, Alistair Moffat, and Justin Zobel. Score standardization for inter-collection comparison of retrieval systems. In Sung-Hyon Myaeng, Douglas W. Oard, Fabrizio Sebastiani, Tat-Seng Chua, and Mun-Kew Leong, editors, *Proc. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 51–58, Singapore, Singapore, July 2008b.

William Webber, Alistair Moffat, Justin Zobel, and Tetsuya Sakai. Precision-at-ten considered redundant. In Sung-Hyon Myaeng, Douglas W. Oard, Fabrizio Sebastiani, Tat-Seng Chua, and Mun-Kew Leong, editors, *Proc. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 695–696, Singapore, Singapore, July 2008c.

Alistair Moffat, William Webber, and Justin Zobel. Strategic system comparisons via targeted relevance judgments. In Charles L. A. Clarke, Norbert Fuhr, Noriko Kando, W. Kraaij, and A. de Vries, editors, *Proc. 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 375–382, Amsterdam, the Netherlands, July 2007a.

William Webber, Alistair Moffat, and Justin Zobel. Score standardization for robust comparison of retrieval systems. In MingFang Wu, Andrew Turpin, and Amanda Spink, editors, *Proc. 12th Australasian Document*

Computing Symposium, pages 1–8, Melbourne, December 2007a.

Alistair Moffat, William Webber, Justin Zobel, and R. Baeza-Yates. A pipelined architecture for distributed text query evaluation. *Information Retrieval*, 10(3):205–231, June 2007b.

Alistair Moffat, William Webber, and Justin Zobel. Load balancing for term-distributed parallel retrieval. In Susan Dumais, E. Efthimiadis, David Hawking, and Kalervo Järvelin, editors, *Proc. 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 348–355, Seattle, Washington, USA, August 2006.

William Webber and Alistair Moffat. In search of reliable retrieval experiments. In J. Kay, Andrew Turpin, and Ross Wilkinson, editors, *Proc. 10th Australasian Document Computing Symposium*, pages 26–33, Sydney, December 2005.

N. Lester, Alistair Moffat, William Webber, and Justin Zobel. Space-limited ranked query evaluation using adaptive pruning. In Anne H.H. Ngu, Masaru Kitsuregawa, Erich J. Neuhold, Jen-Yao Chung, and Quan Z. Sheng, editors, *Proc. 6th International Conference on Web Informations Systems Engineering*, volume 3806 of *Lecture Notes in Computer Science*, pages 470–477, New York, New York, USA, November 2005.

Non-peer-reviewed venues

William Webber, Falk Scholer, MingFang Wu, Xiuzhen Zhang, Douglas W. Oard, Phil Farrelly, Sandra Potter, Steven Dick, and Phill Bertolus. The Melbourne team at the TREC 2010 legal track. In Ellen Voorhees and Lori P. Buckland, editors, *Proc. 19th Text REtrieval Conference*, pages 49:1–12, Gaithersburg, Maryland, USA, November 2010d.

William Webber. Evaluating the effectiveness of keyword search. *IEEE Data Eng. Bull.*, 33(1):54–59, 2010a.

William Webber, V. N. Anh, and Alistair Moffat. The University of Melbourne in the Million Query Track of TREC 2007. In Ellen Voorhees and Lori P. Buckland, editors, *Proc. 16th Text REtrieval Conference*, page 86:1:5, Gaithersburg, Maryland, USA, November 2007b. NIST Special Publication 500-274.

William and Webber and Alistair Moffat. Melbourne University at the 2006 Terabyte Track. In Ellen Voorhees and Lori P. Buckland, editors, *Proc. 15th Text REtrieval Conference*, pages 103:1–8, Gaithersburg, Maryland, USA, November 2006. NIST Special Publication 500-272.

V. N. Anh, William Webber, and Alistair Moffat. Melbourne University 2005: Enterprise and Terabyte tracks. In Ellen Voorhees and Lori P. Buckland, editors, *Proc. 14th Text REtrieval Conference*, pages 114:1–8, Gaithersburg, Maryland, USA, November 2005. NIST Special Publication 500-266.

Yaniv Bernstein, Bodo Billerbeck, Steven Garcia, Nicholas Lester, Falk Scholer, Justin Zobel, and William Webber. RMIT university at TREC 2005: Terabyte and robust track. In Ellen Voorhees and Lori P. Buckland, editors, *Proc. 14th Text REtrieval Conference*, pages 78:1–12, Gaithersburg, Maryland, USA, November 2005. NIST Special Publication 500-266.

Bodo Billerbeck, Adam Cannane, Abhijit Chattaraj, Nicholas Lester, William Webber, Hugh Williams, John Yiannis, and Justin Zobel. RMIT university at TREC 2004. In Ellen Voorhees and Lori P. Buckland, editors, *Proc. 13th Text REtrieval Conference*, pages 55:1–8, Gaithersburg, Maryland, USA, November 2004. NIST Special Publication 500-261.

Theses

William Webber. *Measurement in Information Retrieval Evaluation*. PhD thesis, University of Melbourne, Melbourne, Australia, September 2010b.

William Webber. Design and implementation of a pipelined distributed information retrieval architecture. Master's thesis, University of Melbourne, Melbourne, Australia, June 2007.

PROFESSIONAL ACTIVITIES

Journal editorships

Information Retrieval 2011 to present
Member of Editorial Board

Conference and workshop chairs

SIGIR July, 2013
Area co-chair, Evaluation
36th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland

EVIA June, 2013
Co-chair (with Ruihua Song)
5th International Workshop on Evaluating Information, Tokyo, Japan

EVIA December, 2011
Co-chair (with Mark Sanderson and Kazuaki Kishida)
4th International Workshop on Evaluating Information, Tokyo, Japan

AIRS December, 2010
Area co-chair, Evaluation (with Tetsuya Sakai)
6th Asian Information Retrieval Societies Conference, Taipei, Taiwan

EVIA June, 2010
Co-chair (with Tetsuya Sakai and Mark Sanderson)
3rd International Workshop on Evaluating Information, Tokyo, Japan

Conference and workshop program committees

AIRS December, 2012
Program committee
8th Asia Information Retrieval Societies Conference, Tianjin, China

SIGIR August, 2012
Program committee
35th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, Oregon

CIKM October, 2011
Program committee
20th ACM International Conference on Information and Knowledge Management, Glasgow, United Kingdom

IJCAI July, 2011
Senior program committee
22nd International Joint Conference on Artificial Intelligence

SIGIR July, 2011
Program committee

34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,
Beijing, China

ADCS December, 2010

Program committee

15th Australian Document Computing Symposium, Melbourne, Australia

SIGIR July, 2010

Program committee

33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,
Geneva, Switzerland

ADCS December, 2009

Program committee

14th Australian Document Computing Symposium, Sydney, Australia

CIKM November, 2009

Program committee

18th ACM International Conference on Information and Knowledge Management, Hong Kong, China

ADCS December, 2007

Program committee

12th Australian Document Computing Symposium, Melbourne, Australia

Journal reviewing

ACM Transactions on Information Systems 2007 to present

**Journal of the American Society of Information
Science and Technology** 2007,2012

Information Retrieval (Kluwer) 2007 to present

The Computer Journal 2011

**IEEE Transactions on Knowledge and Data
Engineering** 2011

**Transaction for the Association for Computational
Linguistics** 2013

Research lab affiliations

E-Discovery Lab 2011 to present
University of Maryland, United States

**Computational Linguistics and Information
Processing (CLIP) Lab** 2011 to present
University of Maryland, United States

Industry collaborations

VSCL@TREC 2011
Technical lead
Collaborative effort of the Victorian Society for Computers and the Law at the TREC 2011 Legal Track.

VSCL@TREC 2010
Technical lead

Collaborative effort of the Victorian Society for Computers and the Law at the TREC 2010 Legal Track.

TEACHING AND TALKING

Lecturing

Information Technology Spring Semester, 2012

University of Maryland, United States

Masters level course in Library Science program. Quality of teaching outcome: 3.4/5.0.

Programming the Machine Semester 2, 2010

University of Melbourne, Australia

Second year undergraduate course in assembler and C programming. Quality of teaching outcome: 3.9/5.0.

Text and Document Management Semester 2, 2009

University of Melbourne, Australia

Masters level course in text compression and information retrieval. Quality of teaching outcome: 4.5/5.0.

Introduction to C++ November–December, 2002

RMIT University, Melbourne, Australia

Six-week industry course in C++.

Webinars

Best Practices for Leveraging Predictive Technologies Exterro Inc.

January, 2013

Invited speaker alongside Ralph Losey, Esq. (Partner and National E-Discovery Counsel, Jackson Lewis LLC).

Supervision

Mossaab Bagdouri 2011 to present

University of Maryland, United States

De facto assistant supervisor (with Douglas Oard). PhD candidate, working on evaluation of text classification systems.

Ning Gao 2011 to present

University of Maryland, United States

De facto assistant supervisor (with Douglas Oard). PhD candidate, working on evaluation of IR systems.

Chen Sha 2011

RMIT University, Melbourne, Australia

Co-supervising (with Xiuzhen Zhang) honours thesis, “Active Learning for Biased Training Data”.

Panels

SIGIR 2012 Panel on the Use of Proprietary Data in Information Retrieval Research July, 2012

Panelist

Panelist along with Ricardo Baeza-Yates (Yahoo! Research), Susan Dumais (Microsoft Research), Norbert Fuhr (University of Duisburg-Essen), William Hersh (Origin Health and Science University), and Maarten de Rijke (University of Amsterdam).

RESEARCH INTERESTS

- E-discovery and information retrieval in the legal domain.
- Machine learning for text classification.
- Empirical and statistical methods in computer science research.
- Evaluation of information retrieval system effectiveness.
- Distributed and parallelized information retrieval.
- Time- and space-efficient retrieval technology.

OTHER INTERESTS

- Classical philosophy
- The philosophy and social role of science
- Economic history and critical economic thought
- Spanish early modern and modern poetry