What Would We Like IR Metrics to Measure?

Alistair Moffat
Department of Computing and Information Systems
The University of Melbourne
ammoffat@unimelb.edu.au

Abstract. The field of Information Retrieval has a long-standing tradition of rigorous evaluation, and an expectation that proposals for new mechanisms and techniques will either be evaluated in batch-mode experiments against realistic test collections, with results reported derived from standard tools; or will be evaluated through the use of user studies. This emphasis on evidence, and the desire for verification of proposals, has meant that IR *effectiveness measurement* is an important area studied in its own right. The result has been the development of a complex suite of relevance metrics, each of them with seemingly different behavior. Well-known examples include Precision, Recall, Average Precision, Normalized Discounted Cumulative Gain [3], BPref [2], the Q-Measure [6], Rank-Biased Precision (RBP) [4], and so on. In this presentation the underlying question of what it is that a metric should measure is returned to, with a set of desiderata for usefulness used as a starting point for examining the existing palette of metrics. Recent work that has described a goal-sensitive adaptive metric called INST [1, 5] will then be presented.

Acknowledgment. Much of the work presented in this talk has been carried out and published in collaboration with Peter Bailey (Microsoft), Falk Scholer (RMIT University), and Paul Thomas (CSIRO, Microsoft). University of Melbourne collaborators in earlier work in the area of evaluation have included Justin Zobel and William Webber.

References

[1] P. Bailey, A. Moffat, F. Scholer, and P. Thomas. User variability and IR system evaluation. In *Proc. ACM Conf. on Research and Development in Information Retrieval (SIGIR)*, pages 625–634, 2015.

[2] C. Buckley and E. M. Voorhees. Retrieval evaluation with incomplete information. In *Proc. ACM Conf. on Research and Development in Information Retrieval (SIGIR)*, pages 25–32, 2004.

[3] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Trans. on Information Systems*, 20(4):422–446, 2002.

[4] A. Moffat and J. Zobel. Rank-biased precision for measurement of retrieval effectiveness. *ACM Trans. on Information Systems*, 27(1):2.1–2.27, 2008.

[5] A. Moffat, P. Bailey, F. Scholer, and P. Thomas. INST: An adaptive metric for information retrieval evaluation. In *Proc. Australasian Document Computing Symp. (ADCS)*, pages 5:1–5:4, 2015.

[6] T. Sakai and N. Kando. On information retrieval metrics designed for evaluation with incomplete relevance assessments. *Information Retrieval*, 11(5):447–470, 2008.

Biography. Alistair Moffat has been involved in research in information retrieval for more than 25 years. He has published numerous papers in the areas of index compression, text compression, and dynamic pruning mechanisms, all of which help support efficient ranked querying. Alistair is a co-author of the 1991 (revised 1994) book *Managing Gigabytes*, and also co-author of the 2002 book *Compression and Coding Algorithms*. Much of Alistair's recent work has examined the issue of IR system evaluation, and, with other co-authors in Australia, he has focused on the relationship between models of user interactions with search results pages, and the effectiveness metrics that those interactions correspond to. Alistair was co-Chair for SIGIR 1998 in Melbourne, and for CIKM 2015, also held in Melbourne; and co-Program Committee Chair for SIGIR 2005 (Salvador, Brazil) and SIGIR 2015 (Santiago, Chile).

Alistair has been a teaching/research faculty member at the University of Melbourne for nearly thirty years, and was Department Chair from 2007–2011. During those thirty years he has taught programming skills to well in excess of 10,000 undergraduate students, has authored a popular (he insists that it is popular, because all his students have been told to purchase it) C programming textbook (*Programming, Problem Solving, and Abstraction with C*, 2002, revised 2012), and has received awards for his teaching and lecturing skills.

Alistair's PhD was completed in 1985, at the University of Canterbury, in New Zealand.