Information Retrieval: The Early Years

Other titles in Foundations and Trends® in Information Retrieval

Web Forum Retrieval and Text Analytics: A Survey
Doris Hoogeveen, Li Wang, Timothy Baldwin and Karin M. Verspoor
ISBN: 978-1-68083-350-8

Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting
Jun Wang, Weinan Zhang and Shuai Yuan
ISBN: 978-1-68083-310-2

Applications of Topic Models
Jordan Boyd-Graber, Yuening Hu and David Mimno
ISBN: 978-1-68083-308-9

Searching the Enterprise
Udo Kruschwitz and Charlie Hull
ISBN: 978-1-68083-304-1

Information Retrieval: The Early Years

Donna Harman
National Institute of Standards and Technology, USA
donna.harman@nist.gov

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

D. Harman. *Information Retrieval: The Early Years*. Foundations and Trends® in Information Retrieval, vol. 13, no. 5, pp. 425–577, 2019.

ISBN: 978-1-68083-585-4

Foundations and Trends® in Information Retrieval
Volume 13, Issue 5, 2019

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2019, Volume 13, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Contents

Information Retrieval: The Early Years

Donna Harman

National Institute of Standards and Technology, USA; donna.harman@nist.gov

ABSTRACT

Information retrieval, the science behind search engines, had its birth in the late 1950s. Its forbearers came from library science, mathematics and linguistics, with later input from computer science. The early work dealt with finding better ways to index text, and then using new algorithms to search these (mostly) automatically built indexes. Like all computer applications, however, the theory and ideas were limited by lack of computer power, and additionally by lack of machine-readable text. But each decade saw progress, and by the 1990s, it had flowered. This monograph tells the story of the early history of information retrieval (up until 2000) in a manner that presents the technical context, the research and the early commercialization efforts.

Donna Harman (2019), "Information Retrieval: The Early Years", Foundations and Trends® in Information Retrieval: Vol. 13, No. 5, pp 425–577. DOI: 10.1561/1500000065.

1

Introduction

This monograph traces the evolution of information retrieval, telling a story starting before 1960 and ending in 2000. This evolution was pushed by an ever-growing demand for better ways of finding information, a push that led to new (mostly) government services and to accelerated research. Part of the story is how the early operational systems developed and were able to leverage the minimal technology to produce usable services. Another part is how technology improved, including computer speed, memory, and networking. But the largest part is how the research grew from a small nucleus of ideas drawn from library science, mathematics, and linguistics to the powerhouse that exists today.

I was very privileged to have worked in the SMART lab at Cornell with Prof. Salton in the mid to late 1960s. This included watching Michael Keen do the first in-depth analysis of experimental results on the IRE, ADI, and Cranfield collections. It also included helping to build and analyze the initial MEDLARS test collection, specifically built to compare ranking and Boolean retrieval. They were exciting times–we thought that the new ranking algorithms would shortly revolutionize the retrieval scene.

Alas, when I next got involved in information retrieval in the mid 1980s, Boolean retrieval dominated the commercial world and the core information retrieval research community was very marginalized. This was thankfully about to change as technology improved.

In late 1989 I was asked to build the new large TIPSTER test collection and I initiated the TREC conferences in 1992. The early 1990s brought the excitement of watching the flowering of the ranking techniques in many diverse retrieval models. The arrival of the Internet and the search engines, all using some type of ranking, was indeed a triumph for the retrieval community.

It is clearly impossible to tell the full history of forty plus years in a limited monograph. Instead I have concentrated on the story about how the field got going, starting with the critical issues in indexing and then the early experiments in searching. The story then follows these early themes as they develop. This is set in the context of the available computer technology and the separate but parallel story of the early operational systems, such as MEDLARS.

The emphasis is on how the ideas built on one another, with papers selected that reflect the main experimental themes. My own biases influence the story in that most papers have an experimental flavor as opposed to ones more theoretical in nature.

Chapter 2 deals with the early "automatic" indexing experiments, which laid the groundwork for the field. Chapter 3 (the 1960s) continues with further work on indexing, such as the Cranfield experiments, and then looks at the early experiments with automatic searching. Chapter 4 (the 1970s) introduces the work in probabilistic models, including the many experiments in term weighting, and discusses the expansion of the operational systems. By the 1980s (Chapter 5) the (mostly commercial) operational systems dominate the scene, but research continued, extending the work of the 1970s. Chapter 6 (the 1990s) illustrates the explosion of the research, both in the diversity of topics and the diversity of researchers (hence only a sample of the research is discussed, mostly following the themes from earlier chapters).

References

Aitchison, T. M. and J. M. Tracy (1970), 'Comparative Evaluation of Index Languages: Part II, Results'. Institution of Electrical Engineers.

Allan, J., J. Callan, W. Croft, L. Ballesteros, D. Byrd, R. Swan, and J. Xu (1998a), 'INQUERY Does Battle With TREC-6'. In: *Proceedings of the Sixth Text REtrieval Conference (TREC-6)*. pp. 169–206.

Allan, J., R. Papka, and V. Lavrenko (1998b), 'On-line New Event Detection and Tracking'. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 37–45.

Bagg, T. C. and M. E. Stevens (1961), *Information Selection Systems Retrieval Replica Copies: a State-of-the-Art Report*, Vol. 157. US Dept. of Commerce, National Bureau of Standards.

Barker, F. H., D. C. Veal, and B. K. Wyatt (1972a), 'Comparative Efficiency of Searching Titles, Abstracts, and Index Terms in a Free-Text Data Base'. *Journal of Documentation* **28**(1), 22–36.

Barker, F. H., B. K. Wyatt, and D. C. Veal (1972b), 'Report on the Evaluation of an Experimental Computer-Based Current-Awareness Service for Chemists'. *Journal of the Association for Information Science and Technology* **23**(2), 85–99.

Barry, C. L. (1994), 'User-defined Relevance Criteria: an Exploratory Study'. *Journal of the American Society for Information Science* **45**(3), 149.

Barry, C. L. and L. Schamber (1998), 'Users' Criteria for Relevance Evaluation: A Cross-Situational Comparison'. *Information Processing and Management* **34**(2/3), 219–236.

Bates, M. J. (1989), 'The Design of Browsing and Berrypicking Techniques for the Online Search Interface'. *Online Review* **13**(5), 407–424.

Belew, R. K. (1989), 'Adaptive Information Retrieval: Using a Connectionist Representation to Retrieve and Learn about Documents'. In: *Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 11–20.

Belkin, N. (1980), 'Anomalous States of Knowledge as a Basis for Information Retrieval'. *The Canadian Journal of Information Science* **5**, 133–143.

Belkin, N., R. Oddy, and H. Brooks (1982), 'ASK for Information Retrieval: Part II. Results of a Design Study'. *Journal of Documentation* **38**, 145–164.

Berger, A. and J. Lafferty (1999), 'Information Retrieval as Statistical Translation'. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Blair, D. C. and M. E. Maron (1985), 'An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System'. *Communications of the ACM* **28**(3), 289–299.

Bohnert, L. M. (1955), 'Two Methods of Organizing Technical Information for Search'. *Journal of the Association for Information Science and Technology* **6**(3), 134–151.

Bookstein, A. and D. R. Swanson (1974), 'Probabilistic Models for Automatic Indexing'. *Journal of the Association for Information Science and Technology* **25**(5), 312–316.

Bourne, C. P. (1961), 'The Historical Development and Present State-of-the-Art of Mechanized Information Retrieval Systems'. *Journal of the Association for Information Science and Technology* **12**(2), 108–110.

Bourne, C. P. (1962), 'The World's Technical Journal Literature: An Estimate of Volume, Origin, Language, Field, Indexing, and Abstracting'. *Journal of the Association for Information Science and Technology* **13**(2), 159–168.

Bourne, C. P. and T. B. Hahn (2003), *A History of Online Information Services, 1963–1976*. MIT press.

Bracken, R. H. and B. G. Oldfield (1956), 'A General System for Handling Alphameric Information on the IBM 701 Computer'. *Journal of the ACM (JACM)* **3**(3), 175–180.

Bracken, R. H. and H. Tillitt (1957), 'Information Searching with the 701 Calculator'. *Journal of the ACM (JACM)* **4**(2), 131–136.

Brin, S. and L. Page (1998), 'The Anatomy of a Large-scale Hypertextual Web Search Engine'. In: *Proceedings of the Seventh International Conference on World Wide Web 7*. pp. 107–117.

Brzozowski, J. (1983), 'MASQUERADE: Searching the Full Text of Abstracts using Automatic Indexing'. *Journal of Information Science* **6**(2-3), 67–73.

Buckland, M. K. (1992), 'Emanuel Goldberg, Electronic Document Retrieval, and Vannevar Bush's Memex'. *Journal of the American Society for Information Science (1986–1998)* **43**(4), 284.

Buckley, C. and A. F. Lewit (1985), 'Optimization of Inverted Vector Searches'. In: *Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 97–110.

Buckley, C. and E. M. Voorhees (2000), 'Evaluating Evaluation Measure Stability'. In: *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 33–40.

Burke, C. (1992), 'The Other Memex: The Tangled Career of Vannevar Bush's Information Machine, the Rapid Selector'. *Journal of the American Society for Information Science (1986–1998)* **43**(10), 648.

Bush, V. (1945), 'As We May Think'. *The Atlantic*. pp. 1–26.

Caid, W. R., S. T. Dumais, and S. I. Gallant (1995), 'Learned Vector-Space Models for Document Retrieval'. *Information Processing and Management* **31**(3), 419–429.

Callan, J. P., Z. Lu, and W. B. Croft (1995), 'Searching Distributed Collections with Inference Networks'. In: *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 21–28.

Carbonell, J. and J. Goldstein (1998), 'The Use of MMR, Dversity-Based Reranking for Reordering Documents and Producing Summaries'. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 335–336.

Cerf, V. and R. Kahn (1974), 'A Protocol for Packet Network Intercommunication'. *IEEE Transactions on Communications* **22**(5), 637–648.

Chang, Y., C. Cirillo, and J. Razon (1969), 'Evaluation of Feedback Retrieval using Modified Freezing, Residual Collection and Control Groups'. In: *Scientific Report ISR-16 to NSF*. Cornell University, Ithaca, N.Y., Chapt. X.

Chang, Y., C. Cirillo, and J. Razon (1971), 'Evaluation of Feedback Retrieval using Modified Freezing, Residual Collection and Control Groups'. In: *The SMART Retrieval System*. Prentice-Hall, Englewood Cliffs, New Jersey.

Chuklin, A., I. Markov, and M. d. Rijke (2015), 'Click Models for Web Search'. *Synthesis Lectures on Information Concepts, Retrieval, and Services* **7**(3), 1–115.

Cleverdon, C. (1962), *Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems*. Aslib Cranfield Research Project, Cranfield, U.K.

Cleverdon, C. (1970), *The Effect of Variations in Relevance Assessments in Comparative Experimental Tests of Index Languages*. Cranfield Library Report No. 3, Cranfield, U.K.

Cleverdon, C. (1990), 'Letter to the Editor'. *Online Review* **14**(1), 35.

Cleverdon, C. (1991), 'The Signficance of the Cranfield Tests on Index Languages'. In: *Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 3–12.

Cleverdon, C. and E. Keen (1966), *Factors Determining the Performance of Indexing Systems, Vol. 2: Test Results*. Aslib Cranfield Research Project, Cranfield, U.K.

Cleverdon, C., J. Mills, and E. Keen (1966), *Factors Determining the Performance of Indexing Systems, Vol. 1: Design*. Aslib Cranfield Research Project, Cranfield, U.K.

Codd, E. F. (1970), 'A Relational Model of Data for Large Shared Data Banks'. **13**(6), 377–387.

Cooper, W. (1968), 'Expected Search Length: A Single Measure of Retrieval Effectiveness Based on the Weak Ordering Action of Retrieval Systems'. *American Documentation* **January** pp. 30–41.

Cooper, W. (1971), 'A Definition of Relevance for Information Retrieval'. *Information Storage and Retrieval* **7**, 19–37.

Cooper, W. S. (1983), 'Exploiting the Maximum Entropy Principle to Increase Retrieval Effectiveness'. *Journal of the Association for Information Science and Technology* **34**(1), 31–39.

Croft, W. B. (1981), 'Document Representation in Probabilistic Models of Information Retrieval'. *Journal of the Association for Information Science and Technology* **32**(6), 451–457.

Croft, W. B. (1983), 'Experiments with Representation in a Document-Retrieval System'. *Information Technology-Research Development Applications* **2**(1), 1–21.

Croft, W. B., T. Lucia, and P. R. Cohen (1988), 'Retrieving Documents by Plausible Inference: a Preliminary Study'. In: *Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 481–494.

Croft, W. B. and R. H. Thompson (1987), 'I^{3}R: A New Approach to the Design of Document Retrieval Systems'. *Journal of the American Society for Information Science* **38**(6), 389.

Croft, W. B., H. R. Turtle, and D. D. Lewis (1991), 'The Use of Phrases and Structured Queries in Information Retrieval'. In: *Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 32–45.

Cutting, D. R., D. R. Karger, J. O. Pedersen, and J. W. Tukey (1992), 'Scatter/gather: A Cluster-based Approach to Browsing Large Document Collections'. In: *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 318–329.

Das-Gupta, P. and J. Katzer (1983), 'A Study of the Overlap Among Document Representations'. In: *Proceedings of the 6th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 106–114.

Davenport, L. and B. Cronin (1987), 'Marketing Electronic Information'. *Online Review* **11**(1), 39–47.

Dee, C. R. (2007), 'The Development of the Medical Literature Analysis and Retrieval system (MEDLARS)'. *Journal of the Medical Library Association: JMLA* **95**(4), 416.

Deerwester, S., S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman (1990), 'Indexing by Latent Semantic Analysis'. *Journal of the American Society for Information Science* **41**(6), 391–407.

Dennis, S. F. (1964), 'The Construction of a Thesaurus Automatically from a Sample of Text'. In: *Proceedings of the Symposium on Statistical Association Methods For Mechanized Documentation*. pp. 61–148.

Devlin, B. (1983), 'An Overview of Sense-Making Research: Concepts, Methods and Results'. In: *Proceedings of the Annual Meeting of the International Communication Association*. pp. 1–72.

Diriye, A., R. White, G. Buscher, and S. Dumais (2012), 'Leaving So Soon?: Understanding and Predicting Web Search Abandonment Rationales'. In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. pp. 1025–1034.

Doszkocs, T. E. (1982), 'From Research to Application: the CITE Natural Language System'. In: *Research and Development in Information Retrieval*. pp. 251–262.

Doszkocs, T. E. and B. A. Rapp (1979), 'Searching MEDLINE in English: a Prototype User Interface with Natural Language Query, Ranked Output, and Relevance Feedback'. In: *Proceedings of the 42nd Annual Meeting of the American Society for Information Science*. pp. 131–139.

Edmundson, H. P. and R. E. Wyllys (1961), 'Automatic Abstracting and Indexing Survey and Recommendations'. *Communications of the ACM* **4**(5), 226–234.

Eliot, S. and J. Rose (2009), *A Companion to the History of the Book*, Vol. 98. John Wiley & Sons.

Fidel, R. (1984), 'Online Searching Styles: A Case-Study-Based Model of Searching Behavior'. *Journal of the Association for Information Science and Technology* **35**(4), 211–221.

Fidel, R. (1988), 'Factors Affecting the Selection of Search Keys'. In: *Proceedings of the 51st Annual Meeting of the American Society for Information Science*, Vol. 25, pp. 76–79.

Fidel, R. (1991), 'Searchers' Selection of Search Keys: I. The Selection Routine, II. Controlled Vocabulary or Free-Text Searching, III. Searching Styles'. *Journal of the American Society for Information Science* **42**(7), 490–527.

Fox, E. (1983), *Characteristics of Two New Experimental Collections in Computer and Information Science Containing Textual and Bibliographic Concepts.* Technical Report TR 83-561, Cornell University: Computing Science Department.

Frakes, W. B. and R. Baeza-Yates (1992), *Information Retrieval: Data Structures and Algorithms*. Prentice Hall PTR.

Fuhr, N. (1989), 'Models for Retrieval with Probabilistic Indexing'. *Information Processing and Management* **25**(1), 55–72.

Fuhr, N. and C. Buckley (1991), 'A Probabilistic Learning Approach for Document Indexing'. *ACM Transactions on Information Systems (TOIS)* **9**(3), 223–248.

Fuhr, N., J. Kamps, M. Lalmas, and A. Trotman (2008), *Focused Access to XML Documents: 6th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2007, Revised and Selected Papers*, Vol. 4862. Springer.

Furnas, G. W., S. Deerwester, S. T. Dumais, T. K. Landauer, R. A. Harshman, L. A. Streeter, and K. E. Lochbaum (1988), 'Information Retrieval using a Singular Value Decomposition Model of Latent Semantic Structure'. In: *Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 465–480.

Gilreath, J. and D. L. Wilson (1989), *Thomas Jefferson's Library: a Catalog with the Entries in his Own Order*. The Lawbook Exchange, Ltd.

Giuliano, V. E. and P. E. Jones (1962), 'Linear Associative Information Retrieval'. Technical report, Arthur D. Little, Inc., Cambridge, Massachusetts.

Gull, C. D. (1987), 'Historical Note: Information Science and Technology: From Coordinate Indexing to the Global Brain'. *Journal of the American Society for Information Science* **38**(5), 338–366.

Gull, D. (1956), 'Seven Years of Work on the Organisation of Materials in a Special Library'. *American Documentation* **7**, 320–329.

Hall, H. and N. Weiderman (1967), 'The Evaluation Problem in Rrelevance Feedback'. In: *Scientific Report ISR-12 to NSF*. Cornell University, Ithaca, N.Y., Chapt. XII.

Hancock-Beaulieu, M. and S. Walker (1992), 'An Evaluation of Automatic Query Expansion in an Online Library Catalogue'. *Journal of Documentation* **48**(4), 406–421.

Harman, D. (1988), 'Towards Interactive Query Expansion'. In: *Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 321–331.

Harman, D. (2005), 'The TREC Ad Hoc Experiments'. In: *TREC: Experiment and Evaluation in Information Retrieval*. The MIT Press, Chapt. 4.

Harman, D. (2011), 'Information Retrieval Evaluation'. *Synthesis Lectures on Information Concepts, Retrieval, and Services* **3**(2), 1–119.

Harman, D. and G. Candela (1990), 'Retrieving Records from a Gigabyte of Text on a Minicomputer using Statistical Ranking'. *Journal of the American Society for Information Science* **41**(8), 581.

Harper, D. J. and C. J. van Rijsbergen (1978), 'An Evaluation of Feedback in Document Retrieval using Co-occurrence Data'. *Journal of Documentation* **34**(3), 189–216.

Harter, S. (1971), 'The Cranfield II Relevance Assessments: a Critical Evaluation'. *Library Quarterly* **41**, 229–243.

Harter, S. P. (1975a), 'A Probabilistic Approach to Automatic Keyword Indexing. Part I. On the Distribution of Specialty Words in a Technical Literature'. *Journal of the American Society for Information Science* **26**(4), 197–206.

Harter, S. P. (1975b), 'A Probabilistic Approach to Automatic Keyword Indexing. Part II. On the Distribution of Specialty Words in a Technical Literature'. *Journal of the American Society for Information Science* **26**(5), 280–289.

Hayes, P. J. and S. P. Weinstein (1990), 'CONSTRUE/TIS: A System for Content-Based Indexing of a Database of News Stories'. In: *IAAI*, Vol. 90. pp. 49–64.

Hearst, M., J. Pedersen, P. Pirolli, H. Schutze, G. Grefenstette, and D. Hull (1996), 'Xerox Site Report: Four TREC Tracks'. In: *Proceedings of the Fourth Text REtrieval Conference (TREC-4)*.

Herlocker, J. L., J. A. Konstan, A. Borchers, and J. Riedl (1999), 'An Algorithmic Framework for Performing Collaborative Filtering'. In: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 230–237.

Hersh, W., C. Buckley, T. Leone, and D. Hickam (1994), 'OHSUMED: an Interactive Retrieval Evaluation and New Large Test Collection for Research'. In: *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 192–201.

Holmstrom, J. (1948), 'Section III. Opening Plenary Session'. In: *The Royal Society Scientific Information Conference, 21 June–2 July 1948: Report and papers submitted*.

Ide, E. (1968), 'New Experiments in Relevance Feedback'. In: *Scientific Report ISR-14 to NSF*. Cornell University, Ithaca, N.Y., Chapt. VIII.

Ide, E. (1969), *Relevance Feedback in an Automatic Document Retrieval System*. Scientific Report ISR-15 to NSF.

Ide, E. (1971), 'New Experiments in Relevance Feedback'. In: G. Salton (ed.): *The SMART Retrieval System*. Prentice-Hall, Englewood Cliffs, New Jersey.

Jacobs, P. S. and L. F. Rau (1990), 'SCISOR: Extracting Information from On-Line News'. *Communications of the ACM* **33**(11), 88–97.

Jardine, N. and C. J. van Rijsbergen (1971), 'The Use of Hierarchic Clustering in Information Retrieval'. *Information Storage and Retrieval* **7**(5), 217–240.

Järvelin, K. and J. Kekäläinen (2000), 'IR Evaluation Methods for Retrieving Highly Relevant Documents'. In: *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 41–48.

Kando, N. (1999), 'Proceedings of the First NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition'. In: *Proceedings of the First NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition*.

Kasarda, A. J. and D. J. Hillman (1972), 'The LEADERMART System and Service'. In: *Proceedings of the ACM Annual Conference–Volume 1*. pp. 469–477.

Katzer, J., M. McGill, J. Tessier, W. Frakes, and P. DasGupta (1982), 'Study of the Overlap among Document Representations'. *Information Technology: Research and Development* **1**(2), 261–274.

Keen, E. (1967a), 'Test Environment'. In: *Scientific Report ISR-13 to NSF*. Cornell University, Ithaca, N.Y., Chapt. I.

Keen, E. (1967b), 'Evaluation Parameters'. In: *Scientific Report ISR-13 to NSF*. Cornell University, Ithaca, N.Y., Chapt. II.

Keen, E. (1967c), 'Search Matching Functions'. In: *Scientific Report ISR-13 to NSF*. Cornell University, Ithaca, N.Y., Chapt. III.

Keen, E. (1967d), 'Document Length'. In: *Scientific Report ISR-13 to NSF*. Cornell University, Ithaca, N.Y., Chapt. V.

Keen, E. (1967e), 'Suffix Dictionaries'. In: *Scientific Report ISR-13 to NSF*. Cornell University, Ithaca, N.Y., Chapt. VI.

Keen, E. (1967f), 'Thesaurus, Phrase and Hierarchy Dictionaries'. In: *Scientific Report ISR-13 to NSF*. Cornell University, Ithaca, N.Y., Chapt. VII.

Keen, E. M. (1973), 'The Aberystwyth Index Languages Test'. *Journal of Documentation* **29**(41), 1–35.

Kent, A., M. M. Berry, F. U. Luehrs, and J. W. Perry (1955), 'Machine Literature Searching VIII. Operational Criteria for Designing Information Retrieval Systems'. *Journal of the Association for Information Science and Technology* **6**(2), 93–101.

Kent, A., J. W. Perry, and M. M. Berry (1954), 'Machine Literature Searching V. Definition and Systematization of Terminology for Code Development'. *Journal of the Association for Information Science and Technology* **5**(3), 166–173.

Kilgour, F. G. (1997), 'Origins of Coordinate Searching'. *Journal of the Association for Information Science and Technology* **48**(4), 340–348.

Kleinberg, J. M. (1998), 'Authoritative Sources in a Hyperlinked Environment'. In: *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*.

Kleinberg, J. M. (1999), 'Authoritative Sources in a Hyperlinked Environment'. *Journal of the ACM (JACM)* **46**(5), 604–632.

Knuth, D. (1973), *The Art of Computer Programming, Vol. 3: Sorting and Searching*. Addison-Wesley, Reading, MA.

Koenemann, J. and N. J. Belkin (1996), 'A Case for Interaction: A Study of Interactive Information Retrieval Behavior and Effectiveness'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 205–212.

Kwok, K. L. (1995), 'A Network Approach to Probabilistic Information Retrieval'. *ACM Trans. Inf. Syst.* **13**(3), 324–353.

Lafferty, J. and C. Zhai (2001), 'Document Language Models, Query models, and Risk Minimization for Information Retrieval'. In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 403–410.

Lancaster, F. (1969), *Evaluation of the MEDLARS Demand Search Service*. National Library of Medicine, Washington, D.C.

Lancaster, F., R. L. Rapport, and J. Penry (1972), 'Evaluating the Effectiveness of an On-line, Natural Language Retrieval System'. *Information Storage and Retrieval* **8**(5), 223–245.

Lavrenko, V. and W. B. Croft (2001), 'Relevance Based Language Models'. In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 120–127.

Lee, D. L. (1990), 'Special Issue on Document Retrieval'. *Quarterly Bulletin of the IEEE Computer Society on Data Engineering* **13**(1), 1–63.

Lesk, M. (1995), 'The Seven Ages of Information Retrieval'. UDT Occassional Paper 5, IFLA.

Lesk, M. (2018), Personal communication.

Lesk, M., D. Harman, E. Fox, and C. Buckley (1997), 'The SMART Lab Report'. *SIGIR Forum* pp. 2–22.

Lesk, M. and G. Salton (1968), 'Relevance Assessments and Retrieval System Evaluation'. In: *Scientific Report ISR-14 to NSF*. Cornell University, Ithaca, N.Y., Chapt. III.

Lesk, M. E. (1969), 'Word-Word Associations in Document Retrieval Systems'. *Journal of the American Society for Information Science* **20**(1), 27–38.

Lesk, M. E. and G. Salton (1969), 'Relevance Assessments and Retrieval System Evaluation'. *Information Storage and Retrieval* **4**, 343–359.

Lewis, D. D. and W. A. Gale (1994), 'A Sequential Algorithm for Training Text Classifiers'. In: *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 3–12.

Lovins, J. B. (1968), 'Development of a Stemming Algorithm'. *Mechanical Translation and Computational Linguistics* **11**(1 and 2), 22–31.

Luhn, H. (1957), 'A Statistical Approach to Mechanized Encoding and Searching of Literary Information'. *IBM Journal* **October**, 309–317.

Luhn, H. P. (1953), 'A New Method of Recording and Searching Information'. *American Documentation* **4**(1), 14–16.

Luhn, H. P. (1958a), 'The Automatic Creation of Literature Abstracts'. *IBM Journal of Research and Development* **2**(2), 159–165.

Luhn, H. P. (1958b), 'A Business Intelligence System'. *IBM Journal of Research and Development* **2**(4), 314–319.

Luhn, H. P. (1960), 'Key Word-in-Context Index for Technical Literature (KWIC Index)'. *American Documentation* **11**(4), 288–295.

Luhn, H. P. (1961), 'Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment'. *Journal of the Association for Information Science and Technology* **12**(2), 131–138.

Maron, M. and J. Kuhns (1960), 'On Relevance, Probabilistic Indexing and Information Retrieval'. *Journal of the ACM* **7**, 216–244.

Maron, M., J. Kuhns, and L. Ray (1959), 'Probabilistic Indexing. a Statistical Technique for Document Identification and Retrieval'. Technical report, Thomson Ramo Woolridge, Inc., Los Angeles, CA.

Maron, M. E. (1961), 'Automatic Indexing: an Experimental Inquiry'. *Journal of the ACM (JACM)* **8**(3), 404–417.

Maron, M. E. (2008), 'An Historical Note on the Origins of Probabilistic Indexing'. *Information Processing & Management* **44**(2), 971–972.

McCarn, D. B. (1971), 'Networks with Emphasis on Planning an On-Line Bibliographic Access System'. *Information Storage and Retrieval* **7**(6), 271–279.

McCarn, D. B. (1978), 'Online Systems-Techniques and Services'. *Annual Review of Information Science and Technology* **13**, 85–124.

McGill, M. J., L. C. Smith, S. Davidson, and T. Noreault (1976), 'Syracuse Information Retrieval Experiment (SIRE): Design of an On-line Bibliographic Retrieval System'. *SIGIR Forum* **10**(4), 37–44.

Miller, W. L. (1971), 'A Probabilistic Search Strategy for MEDLARS'. *Journal of Documentation* **27**(4), 254–266.

Mitev, N., G. Venner, and S. Walker (1985), *Designing an Online Public Acess Catalogue: Okapi, a Catalogue on a Local Area Network*. Library and Information Research Report 39, London: British Library.

Mooers, C. N. (1950), 'Information Retrieval Viewed as Temporal Signaling'. In: *Proceedings of the International Congress of Mathematicians*, Vol. 1. pp. 572–573.

Mooers, C. N. (1951), 'Zatocoding applied to mechanical organization of knowledge'. *Journal of the Association for Information Science and Technology* **2**(1), 20–32.

Mooers, C. N. (1960), 'The Next Twenty Years in Information Retrieval; Some Goals and Predictions'. *Journal of the Association for Information Science and Technology* **11**(3), 229–236.

Mothe, J. (1994), 'Search Mechanisms Using a Neural Network Model: Comparison with the Vector Space Model'. In: *Intelligent Multimedia Information Retrieval Systems and Management – Volume 1*. pp. 275–294.

Needham, E. and K. Spärck Jones (1964), 'KEYWORDS AND CLUMPS: Recent work on Information Retrieval at the Cambridge Language Research Unit'. *Journal of Documentation* **20**(1), 5–15.

Oddy, R. (1977), 'Information Retrieval through Man and Machine Dialog'. *Journal of Documentation* **33**(1), 1–14.

Page, L., S. Brin, R. Motwani, and T. Winograd (1998), *The PageRank Citation Ranking: Bringing Order to the Web.*

Pejtersen, A. M. (1989), 'A Library System for Information Retrieval based on a Cognitive Task Analysis and Supported by an Icon-Based Interface'. In: *Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 40–47.

Peters, C. (2003), *Cross-Language Information Retrieval and Evaluation: Workshop of Cross-Language Evaluation Forum, CLEF 2000, Lisbon, Portugal, September 21–22, 2000, Revised Papers*, Vol. 2069.

Ponte, J. M. and W. B. Croft (1998), 'A Language Modeling Approach to Information Retrieval'. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 275–281.

Porter, M. F. (1980), 'An Algorithm for Suffix Stripping'. *Program* **14**(3), 130–137.

Porter, M. F. (1982), 'Implementing a Probabilistic Information Retrieval System'. *Information Technology: Research and Development* **1**(2), 131–156.

Rath, G., A. Resnick, and T. Savage (1961), 'Comparisons of Four Types of Lexical Indicators of Content'. *Journal of the Association for Information Science and Technology* **12**(2), 126–130.

Rau, L. F. (1987), 'Knowledge Organization and Access in a Conceptual Information System'. *Information Processing and Management* **23**(4), 269 – 283.

Rees, A. (1967), 'Evaluation of Information Systems and Services'. In: *Annual Review of Information Science and Technology*. Interscience, Chapt. 3.

Rees, J. and A. Kent (1958), 'Mechanized Searching Experiments using the WRU Searching Selector'. *Journal of the Association for Information Science and Technology* **9**(4), 277–303.

Rice, B. A. (1985), 'Evaluation of Online Databases and Their Uses in Collections'. *Library Trends* **33**(3), 297–325.

Robertson, S. (1969a), 'The Parametric Description of Retrieval Tests: Part I: The Basic Parameters'. *Journal of Documentation* **25**(1), 1–27.

Robertson, S. (1969b), 'The Parametric Description of Retrieval Tests: Part II: Overall Measures'. *Journal of Documentation* **25**(2), 93–103.

Robertson, S. (2003), 'The Unified Model Revisited'. In: *SIGIR 2003 Workshop on Mathematical/Formal Models in Information Retrieval*.

Robertson, S. and J. Callan (2005), 'Routing and Filtering'. In: *TREC: Experiment and Evaluation in Information Retrieval*. The MIT Press, Chapt. 5.

Robertson, S. and K. Spärck Jones (1976), 'Relevance Weighting of Search Terms'. *Journal of the American Society for Information Science* **27**(3), 129–146.

Robertson, S. E. (1977a), 'The Probability Ranking Principle in IR'. *Journal of Documentation* **33**(4), 294–304.

Robertson, S. E. (1977b), 'Theories and Models in Information Retrieval'. *Journal of Documentation* **33**(2), 126–148.

Robertson, S. E. and M. M. Hancock-Beaulieu (1992), 'On the Evaluation of IR Systems'. *Information Processing and Management* **28**(4), 457–466.

Robertson, S. E., M. Maron, and W. S. Cooper (1982), 'The Unified Probabilistic Model for IR'. In: *Research and Development in Information Retrieval*. pp. 108–117.

Robertson, S. E., C. J. van Rijsbergen, and M. F. Porter (1980), 'Probabilistic Models of Indexing and Searching'. In: *Proceedings of the 3rd Annual ACM Conference on Research and Development in Information Retrieval*. pp. 35–56.

Robertson, S. E. and S. Walker (1994), 'Some Simple Effective Approximations to the 2-poisson Model for Probabilistic Weighted Retrieval'. In: *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 232–241.

Rocchio, J. (1965), 'Relevance Feedback in Information Retrieval'. In: *Scientific Report ISR-9 to NSF*. Harvard University, Cambridge, Massachusetts, Chapt. XXIII.

Rocchio, J. J. (1966), *Document Retrieval Systems – Optimization and Evaluation*. Scientific Report ISR-10 to NSF, Cambridge, Massachusetts.

Rocchio, J. J. (1971), 'Relevance Feedback in Information Retrieval'. In: G. Salton (ed.): *The SMART Retrieval System*. Prentice-Hall, Englewood Cliffs, New Jersey.

Salton, G. (1963), 'Associative Document Retrieval Techniques Using Bibliographic Information'. *Journal of ACM* **10**(4), 440–457.

Salton, G. (1964a), 'The Evaluation of Automatic Retrieval Procedures– selected test results using the SMART system'. In: *Scientific Report ISR-8 to NSF*. Harvard University, Cambridge, Massachusetts, Chapt. IV.

Salton, G. (1964b), 'A Flexible Automatic System for the Organization, Storage and Retrieval of Language Data'. In: *Scientific Report ISR-5 to NSF*. Harvard University, Cambridge, Massachusetts, Chapt. I.

Salton, G. (1964c), *Scientific Report ISR-7 to NSF*. Harvard University, Cambridge, Massachusetts.

Salton, G. (1965), 'The Evaluation of Automatic Retrieval Procedures– Selected Test Results Using the SMART System'. *American Documentation* **16**(3), 209–222.

Salton, G. (1965), *Scientific Report ISR-9 to NSF*. Harvard University, Cambridge, Massachusetts.

Salton, G. (1967), *Scientific Report ISR-9 to NSF*. Harvard University, Cambridge, Massachusetts.

Salton, G. (ed.) (1968), *Automatic Information Organization and Retrieval*. McGraw-Hill Book Co., New York, N.Y.

Salton, G. (1969a), 'Automatic Processing of Foreign Language Documents'. In: *Proceedings of the 1969 Conference on Computational Linguistics*. pp. 1–28, Association for Computational Linguistics.

Salton, G. (1969b), 'A Comparison between Manual and Automatic Indexing Methods'. *Journal of the Association for Information Science and Technology* **20**(1), 61–71.

Salton, G. (ed.) (1971), *The SMART Retrieval System*. Prentice-Hall, Englewood Cliffs, New Jersey.

Salton, G. and C. Buckley (1990), 'Improving Retrieval Performance by Relevance Feedback'. *Journal of the American Society for Information Science* **41**(4), 288–297.

Salton, G., E. A. Fox, and H. Wu (1983), 'Extended Boolean Information Retrieval'. *Communications of the ACM* **26**(11), 1022–1036.

Salton, G. and M. E. Lesk (1965), 'The SMART Automatic Document Retrieval Systems: an Illustration'. *Communications of the ACM* **8**(6), 391–398.

Salton, G. and M. E. Lesk (1968), 'Computer Evaluation of Indexing and Text Processing'. *Journal of the ACM (JACM)* **15**(1), 8–36.

Salton, G. and M. McGill (eds.) (1983), *Introduction to Modern Information Retrieval*. McGraw-Hill Book Co., New York, NY.

Salton, G. and D. Williamson (1968), 'A Comparison Between Manual and Automatic Indexing Methods'. In: *Scientific Report ISR-14 to NSF*. Cornell University, Ithaca, N.Y., Chapt. VI.

Salton, G. and C. Yang (1973), 'On the Specification of Term Values in Automatic Indexing'. *Journal of Documentation* **29**(4), 351–372.

Salton, G., C.-S. Yang, and C. T. Yu (1974), 'A Theory of Term Importance in Automatic Text Analysis'. In: *Scientific Report ISR-22 to NSF*. Cornell University, Ithaca, N.Y., Chapt. III.

Salton, G., C.-S. Yang, and C. T. Yu (1975), 'A Theory of Term Importance in Automatic Text Analysis'. *Journal of the Association for Information Science and Technology* **26**(1), 33–44.

Sanderson, M. (2010), 'Test Collection Based Evaluation of Information Retrieval Systems'. *Foundations and Trends in Information Retrieval* **4**, 247–375.

Sanderson, M. and W. B. Croft (2012), 'The History of Information Retrieval Research'. *Proceedings of the IEEE* **100**(Special Centennial Issue), 1444–1451.

Saracevic, T. (1968), 'Linking Research and Teaching'. *American Documentation* **October**, pp. 398–403.

Saracevic, T. (1971), 'Selected Results from an Inquiry into Testing of Information Retrieval Systems'. *Journal of the American Society for Information Science* **March–April**, pp. 126–139.

Saracevic, T. (2007a), 'Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part II: Nature and Manifestations of Relevance'. *Journal of the American Society for Information Science* **58**(13), 1915–1933.

Saracevic, T. (2007b), 'Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part III: Behavior and Effects of Relevance'. *Journal of the American Society for Information Science* **58**(13), 2126–2144.

Saracevic, T. and P. Kantor (1988a), 'A Study of Information Seeking and Retrieving. II: Searchers, Searches, and Overlap'. *Journal of the American Society for Information Science* **39**(2), 197–216.

Saracevic, T. and P. Kantor (1988b), 'A Study of Information Seeking and Retrieving. II: Users, Questions, and Effectiveness'. *Journal of the American Society for Information Science* **39**(2), 177–196.

Saracevic, T., P. Kantor, A. Y. Chamis, and D. Trivison (1988), 'A Study of Information Seeking and Retrieving. I: Background and Methodology'. *Journal of the American Society for Information Science* **39**(2), 161–176.

Segesta, J. and K. Reid-Green (2002), 'Harley Tillitt and Computerized Library Searching'. *IEEE Annals of the History of Computing* **24**(3), 23–34.

Seymour, T., D. Frantsvog, and S. Kumar (2011), 'History of Search Engines'. *International Journal of Management & Information Systems* **15**(4), 47–58.

Shera, J. H. (1955), 'The Truth, the Whole Truth...'. *American Documentation* **6**, 56.

Siegel, E., K. Kameen, S. Sinn, and F. O. Weise (1984), 'Research Strategy and Methods used to Conduct a Comparative Evaluation of Two Prototype Online Catalog Systems'. In: *Proceedings of the National Online Meeting*. pp. 503–511.

Singhal, A., C. Buckley, and M. Mitra (1996), 'Pivoted Document Length Normalization'. In: *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 21–29.

Smeaton, A. F., P. Over, and W. Kraaij (2006), 'Evaluation Campaigns and TRECvid'. In: *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*. pp. 321–330.

Smith, E. S. (1993), 'On the Shoulders of Giants: From Boole to Shannon to Taube; The Origins and Development of Computerized Information from the Mid-19th Century to the Present'. *Information Technology and Libraries* **12**(2), 217.

Spärck Jones, K. (1972), 'A Statistical Interpretation of Term Specificity and its Application in Retrieval'. *Journal of Documentation* **60**(5), 493–502.

Spärck Jones, K. (1973a), 'Collection Properties Influencing Automatic Term Classifications Performance'. *Information Storage and Retrieval* **9**, 499–513.

Spärck Jones, K. (1973b), 'Index Term Weighting'. *Information Storage and Retrieval* **9**(11), 619–633.

Spärck Jones, K. (1975), 'A Performance Yardstick for Test Collections'. *Journal of Documentation* **31**(4), 266–272.

Spärck Jones, K. (1979a), 'Experiments in Relevance Weighting of Search Terms'. *Information Processing and Management* **15**(3), 133–144.

Spärck Jones, K. (1979b), 'Search Term Relevance Weighting Given Little Relevance Information'. *Journal of Documentation* **35**(1), 30–48.

Spärck Jones, K. (ed.) (1981), *Information Retrieval Experiment*. Butterworths.

Spärck Jones, K. (1988), 'A Look Back and a Look Forward'. In: *Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 13–29.

Spärck Jones, K. (2000), 'Further Reflections on TREC'. *Information Processing and Management* **36**(1), 37–86.

Spärck Jones, K. (2004), 'IDF Term Weighting and IR Research Lessons'. *Journal of documentation* **60**(5), 521–523.

Spärck Jones, K. and D. Jackson (1970), 'The Use of Automatically-Obtained Keyword Classifications for Information Retrieval'. *Information Storage and Retrieval* **5**, 175–201.

Spärck Jones, K. and R. M. Needham (1968), 'Automatic Term Classifications and Retrieval'. *Information Storage and Retrieval* **4**(2), 91–100.

Spärck Jones, K. and C. van Rijsbergen (1976), 'Information Retrieval Test Collections'. *Journal of Documentation* **32**(1), 59–75.

Stanfill, C. and B. Kahle (1986), 'Parallel Free-Text Search on the Connection Machine System'. *Communications of the ACM* **29**(12), 1229–1239.

Stanfill, C. and D. Waltz (1986), 'Toward Memory-based Reasoning'. *Communications of the ACM* **29**(12), 1213–1228.

Stevens, M. E. (1970), *Automatic Indexing: A State-of-the-Art Report.*

Stevens, M. E. and V. E. Giuliano (1965), *Statistical Association Methods for Mechanized Documentation: Symposium Proceedings, Washington, 1964*, Vol. 269. US Government Printing Office.

Stiles, H. E. (1961), 'The Association Factor in Information Retrieval'. *Journal of the ACM (JACM)* **8**(2), 271–279.

Summit, R. K. (1989), 'In Search of the Elusive End User'. *Online Review* **13**(6), 485–491.

Swanson, D. (1971), 'Some Unexplained Aspects of the Cranfield Tests of Indexing Language Performance'. *Library Quarterly* **41**, 223–228.

Swanson, D. R. (1960), 'Searching Natural Language Text by Computer'. *Science* **132**(3434), 1099–1104.

Swets, J. A. (1969), 'Effectiveness of Information Retrieval Methods'. *American Documentation* **January**, pp. 72–89.

Taube, M. and Associates (1955), 'Storage and Retrieval of Information by Means of the Association of Ideas'. *Journal of the Association for Information Science and Technology* **6**(1), 1–18.

Taube, M., C. Gull, and I. S. Wachtel (1952), 'Unit terms in coordinate indexing'. *Journal of the Association for Information Science and Technology* **3**(4), 213–218.

Tenopir, C. (1984), 'Full-text Databases'. *Annual Review of Information Science and Technology* **19**, 215–246.

Turtle, H. (1994), 'Natural Language vs. Boolean Query Evaluation: A Comparison of Retrieval Performance'. In: *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 212–220.

Turtle, H. and W. B. Croft (1989), 'Inference Networks for Document Retrieval'. In: *Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 1–24.

Turtle, H. and W. B. Croft (1991a), 'Evaluation of an Inference Network-Based Retrieval Model'. *ACM Transactions on Information Systems (TOIS)* **9**(3), 187–222.

Turtle, H. R. and W. B. Croft (1991b), 'Efficient Probabilistic Inference for Text Retrieval'. In: *Intelligent Text and Image Handling – Volume 2*. pp. 644–661.

US Public Health Service *et al.* (1963), 'The MEDLARS Story at the National Library of Medicine'. *Washington, D.C., Government Printing Office.*

van Rijsbergen, C. J. (1973), 'Further Experiments with Hierarchic Clustering in Document Retrieval'. *Information Storage and Retrieval* **10**(1), 1–14.

van Rijsbergen, C. J. (1975), *Information Retrieval.* Butterworths.

van Rijsbergen, C. J. (1977), 'A Theoretical Basis for the Use of Co-occurrence Data in Information Retrieval'. *Journal of Documentation* **33**(2), 106–119.

van Rijsbergen, C. J. (1986), 'A New Theoretical Framework for Information Retrieval'. In: *Proceedings of the 9th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* pp. 194–200.

Varlejs, J. (1999), 'Ralph Shaw and the Rapid Selector'. In: *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems.* pp. 148–155.

Vaswani, P. and J. Cameron (1970), *The National Physical Laboratory Experiments in Statistical Word Associations and their Use in Document Indexing and Retrieval.* Publication 42, Division of Computer Science, National Physical Laboratory, Teddington.

Voorhees, E. and D. Harman (eds.) (2005), *TREC: Experiment and Evaluation in Information Retrieval.* The MIT Press.

Voorhees, E. M. (1985), 'The Cluster Hypothesis Revisited'. In: *Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* pp. 188–196.

Voorhees, E. M. (1998), 'Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness'. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* pp. 315–323.

Walker, S. and R. de Vere (1989), *Improving Subject Retrieval in Online Catalogues: 2. Relevance Feedback and Query Expansion.* British Library Research Paper 72, London: British Library.

Walker, S. and M. Hancock-Beaulieu (1991), *Okapi at City, an Evaluation Facility for Interactive IR.* British Library Research Report 6056, London: British Library.

Walker, S. and R. M. Jones (1987), *Improving Subject Retrieval in Online Catalogues: 1. Stemming, Automatic Spelling Correction and Cross-Reference Tables.* British Library Research Paper 24, London: British Library.

Willett, P. (2006), 'The Porter Stemming Algorithm: Then and Now'. *Program* **40**(3), 219–223.

Williams, M. E. (1985a), 'Electronic Databases'. *Science* **228**(4698), 445–450.

Williams, M. E. (1985b), 'Usage and Revenue Data for the Online Database Industry'. *Online Review* **9**(3), 205–210.

Williams, M. E. (1986), 'Online Government Databases–An Analysis'. *Online Review* **10**(4), 227–36.

Williamson, D. (1968), 'A Cornell Implementation of the SMART System'. In: *Scientific Report ISR-14 to NSF*. Cornell University, Ithaca, N.Y., Chapt. VII.

Wolfe, G. (1994), 'The Second Phase of the Revolution has Begun'. *Wired* **2**(10), 116–121.

Xu, J. and W. B. Croft (1996), 'Query Expansion Using Local and Global Document Analysis'. In: *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 4–11.

Zhai, C. and J. Lafferty (2001), 'A Study of Smoothing Methods for Language Models Applied to Information Retrieval'. In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 334–342.

Zobel, J. (1998), 'How Reliable are the Results of Large-Scale Information Retrieval Experiments'. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 307–314.