Signature file:

a text index based on storing a signature for each text block to be able to filter out some blocks quickly.

SIMD:

a parallel computer architecture consisting of a single instruction stream and multiple data streams.

SISD:

a traditional sequential computer architecture consisting of a single instruction stream and a single data stream.

SMP (Symmetric Multiprocessor):

a common MIMD computer architecture where all of the parallel processors have equal access to the same system resources, and any processor can execute any task. The processors may operate independently, or they may cooperate to execute a parallel task concurrently.

Source selection:

the process of selecting one or more document collections from a set of document collections where the selected collections are most likely to contain documents relevant to the query. Commonly used in distributed IR systems.

SQL3:

the upcoming SQL standard for relational databases.

Spatial Access Methods:

file structures for fast storage and retrieval of spatial objects like points, lines, polygons.

Speedup:

a measure of parallel algorithm performance, defined as:

$$S = \frac{\text{Running time of best available sequential algorithm}}{\text{Running time of parallel algorithm}}$$

STARTS (Stanford Proposal for Internet Meta-Searching):

a protocol for distributed, heterogeneous search developed at Stanford University in cooperation with a consortium of search product and search service vendors.

Statistical text compression:

a reference to techniques and methods of text compression based on probability estimates for the rates of occurrence of each symbol.

Stemming:

a technique for reducing words to their grammatical roots.

Stopwords:

words which occur frequently in the text of a document. Examples of stopwords are articles, prepositions, and conjunctions.

Suffix tree and suffix array:

text indices based on a lexicographical arrangement of all the text suffixes.

Syntax tree:

structural interpretation of a query, where the nodes are the operators and the subtrees are the operands.

Tag:

a string which is used to mark the beginning or ending of structural elements in the text.

Term-to-term independence:

a fundamental assumption underlying the classic vector model which states that the vectors used for representing index terms form an orthonormal basis. Such an assumption is usually interpreted to mean pairwise orthogonality (but not in the generalized vector space model).

Text structure:

information present in a text apart from its content, which relates its different portions in a semantically meaningful way.

Textual Image:

image that is mostly text and then can be compressed more than conventional images. In addition, by using OCR, keywords can be extracted and used for retrieving the image.

Thesaurus:

a data structure composed of (1) a pre-compiled list of important words in a given domain of knowledge and (2) for each word in this list, a list of related (synonym) words.

TREC collection:

a reference collection which contains over a million documents and which has been used extensively in the TREC conferences. The TREC collection has been organized by NIST and is becoming a standard for comparing IR models and algorithms.

User information need:

a natural language declaration of the informational need of a user. For instance, '*find documents which discuss the political implications of the Monica Lewinsky scandal in the results of the 1998 elections for the U.S. Congress*'.

User relevance feedback:

an interactive process of obtaining information from the user about the relevance and the non-relevance of retrieved documents.

URL:

stands for Uniform Resource Locator and is used to address any Internet resource, including Web pages.

Vector model:

a classic model of document retrieval based on representing documents and queries as vectors of index terms. The model adopts as its foundation the notion of term-to-term independence.

Vocabulary:

the set of all the different words in a text.

WAIS:

(Wide Area Information Service): suite of protocols designed to publish information and to allow querying of databases in the Internet.

Working memory load:

information about the decisions and choices made during an interactive user session. A system with a well designed interface should keep this information around to facilitate the task of the user (for instance, allowing the user to easily resume temporarily abandoned strategies).

XML:

a subset of SGML, defined for the Web. In XML it is easier to define new markup languages.

Z39.50:

a national standard that has become an international standard. It is a protocol for client/server communication with retrieval systems, allowing a single client to interact with one or more systems, or retrieval systems (or gateways) to communicate with other retrieval systems. It supports connection-oriented sessions, having a system EXPLAIN itself, submission of queries, obtaining information, the results lists, and obtaining retrieved documents.

Zipf's Law:

an empirical rule that describes the frequency of the text words. It states that the i-th most frequent word appears as many times as the most frequent one divided by i^{θ}, for some $\theta \geq 1$.

Zooming (see also *panning*):

action of a movie camera that can either move in for a close-up or back away to get a wider view. This effect can be simulated in a screen window even in the absence of the camera.

References

[1] Ijsbrand Jan Aalbersberg. Incremental relevance feedback. In *Proc. of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 11–22, Copenhagen, Denmark, 1992.

[2] Serge Abiteboul. Querying semi-structured data. In Foto N. Afrati and Phokion Kolaitis, editors, *Int. Conf. on Database Theory (ICDT)*, number 1186 in LNCS, pages 1–18, Delphi, Greece, 1997. Springer-Verlag.

[3] Marc Abrams, editor. *World Wide Web: Beyond the Basics*. Prentice Hall, 1998.

[4] Brent Agnew, Christos Faloutsos, Zhengyu Wang, Don Welch, and Xiaogang Xue. Multimedia indexing over the Web. In *Proc. of the SPIE Conference on Storage and Retrieval for Image and Video Databases*, pages 72–83, San Jose, CA, Feb 1997.

[5] M. Agosti, G. Gradenigo, and P. G. Marchetti. A hypertext environment for interacting with large textual databases. *Information Processing & Management*, 28(3):371–387, 1992.

[6] Rakesh Agrawal, Christos Faloutsos, and Arun Swami. Efficient similarity search in sequence databases. In *Fourth Int. Conf. on Foundations of Data Organization and Algorithms*, pages 69–84, Evanston, Illinois, Oct 1993. ftp://olympos.cs.umd.edu/pub/TechReports/fodo.ps.

[7] Rakesh Agrawal, King-Ip Lin, Harpreet S. Sawhney, and Kyuseok Shim. Fast similarity search in the presence of noise, scaling and translation in time-series databases. In *Proc. of VLDB Conf.*, pages 490–501, Zurich, Switzerland, September 1995.

[8] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. In *Proc. of VLDB Conf.*, pages 487–499, Santiago, Chile, Sept 1994.

[9] A. Aho and M. Corasick. Efficient string matching: an aid to bibliographic search. *Communications of the ACM*, 18(6):333–340, June 1975.

[10] Alexa: Main Page. http://www.alexa.com, 1998.

[11] Alis Technologies, Web languages hit parade. http://babel.alis.com:8080/palmares.html, 1997.

[12] James Allan. Relevance feedback with too much data. In *Proc. of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 337–343, Seattle, WA, USA, 1995.

[13] Bryce L. Allen. *Information Tasks: Toward a User-Centered Approach to Information Systems*. Academic Press, San Diego, CA, 1996.

[14] Robert B. Allen, Pascal Obry, and Michael Littman. An interface for navigating clustered document sets returned by queries. In *Proc. of ACM COCS: Conference on Organizational Computing Systems*, pages 66–171, Milpitis, CA, November 1993.

[15] O. Alonso and R. Baeza-Yates. A model for visualizing large answers in WWW. In *XVIII Int. Conf. of the Chilean CS Society*, pages 2–7, Antofagasta, Chile, 1998. IEEE CS Press.

[16] T. M. Alrashid, J. A. Barker, B. S. Christian, S. C. Cox, M. W. Rabne, E. A. Slotta, and L. R. Upthegrove. Safeguarding Copyrighted Contents, Digital Libraries and Intellectual Property Management, CWRU's Rights Management System. *D-Lib Magazine*, April 1998. http://www.dlib.org/dlib/april98/04barker.html.

[17] AltaVista: Main page. http://www.altavista.com, 1996.

[18] G. M. Amdahl. Validity of the single-processor approach to achieving large scale computing capabilities. In *Proc. AFIPS 1967 Spring Joint Computer Conf.*, volume 30, pages 483–485, Atlantic City, N.J., April 1967.

[19] A. Amir, G. Benson, and M. Farach. Let sleeping files lie: pattern matching in z-compressed files. *Journal of Computer and Systems Sciences*, 52(2):299–307, 1996.

[20] Keith Andrews. Visualizing cyberspace: Information visualization in the Harmony Internet browser. In *Proceedings '95 Information Visualization*, pages 97–104, Atlanta, USA, October 1995.

[21] P. Anick, J. Brennan, R. Flynn, D. Hanssen, B. Alvey, and J. Robbins. A direct manipulation interface for Boolean information retrieval via natural language query. In *Proc. of the 13th Annual International ACM/SIGIR Conference*, pages 135–150, Brussels, Belgium, 1990.

[22] Peter G. Anick. Adapting a full-text information retrieval system to the computer troubleshooting domain. In *Proc. of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 349–358, 1994.

[23] ANSI/NISO Standards, Z39.50-Information Retrieval: Application Service Definition and Protocol Specification, 1995. See http://lcweb.loc.gov/z3950/agency.

[24] P. Apers, H. Blanken, and M. Houtsma. *Multimedia Databases in Perspective.* Springer-Verlag, 1997.

[25] A. Apostolico and Z. Galil. *Combinatorial Algorithms on Words.* Springer-Verlag, New York, 1985.

[26] M. Araújo, G. Navarro, and N. Ziviani. Large text searching allowing errors. In *Proc. WSP'97*, pages 2–20, Valparaíso, Chile, 1997. Carleton University Press.

[27] Y. Arens, C. Chee, C. Hsu, and C. Knoblock. Retrieving and integrating data from multiple information sources. *Journal on Intelligent and Cooperative Information Systems*, 2(2):127–158, 1993.

[28] W. Y. Arms. Key concepts in the architecture of the digital library. *D-Lib Magazine*, July 1995. http://www.dlib.org/dlib/July95/07arms.html.

[29] W. Y. Arms. Relaxing assumptions about the future of digital libraries: The hare and the tortoise. *D-Lib Magazine*, April 1997. http://www.dlib.org/dlib/april97/04arms.html.

[30] W. Y. Arms. Implementing Policies for Access Management. *D-Lib Magazine*, February 1998. http://www.dlib.org/dlib/february98/arms/02arms.html.

[31] W. Y. Arms, C. Blanchi, and E. A. Overly. An Architecture for Information in digital libraries. *D-Lib Magazine*, February 1997. http://www.dlib.org/dlib/february97/cnri/02arms1.html.

[32] Gustavo Arocena and Alberto Mendelzon. WebOQL: Restructuring documents, databases and Webs. In *Int. Conf. on Data Engineering*, pages 24–33, Orlando Florida, 1998.

[33] Gustavo O. Arocena, Alberto O. Mendelzon, and George A. Mihaila. Applications of a Web query language. In *6th Int. WWW Conference*, Santa Clara, CA, USA, April 1997.

[34] AskJeeves: Main Page. http://www.askjeeves.com, 1998.

[35] R. Attar and A. S. Fraenkel. Local feedback in full-text retrieval systems. *Journal of the ACM*, 24(3):397–417, 1977.

[36] R. Baeza-Yates. An extended model for full-text databases. *Journal of Brazilian CS Society*, 3(2):57–64, April 1996.

[37] R. Baeza-Yates, E. Barbosa, and N. Ziviani. Hierarchies of indices for text searching. In *Proceedings of RIAO'94 Intelligent Multimedia Information Retrieval Systems and Management*, pages 11–13, 1994.

[38] R. Baeza-Yates and G. Gonnet. Efficient Text Searching of Regular Expressions. In G. Ausiello, M. Dezani-Ciancaglini, and S. Ronchi Della Rocca, editors, *ICALP'89*, number 372 in LNCS, pages 46–62, Stresa, Italy, 1989. Springer-Verlag.

[39] R. Baeza-Yates and G. Gonnet. A new approach to text searching. *Communications of the ACM*, 35(10):74–82, October 1992.

[40] R. Baeza-Yates and G. Gonnet. Fast text searching for regular expressions or automaton searching on a trie. *J. of the ACM*, 43(6):915–936, Nov 1996.

[41] R. Baeza-Yates and G. Navarro. Integrating contents and structure in text retrieval. *ACM SIGMOD Record*, 25(1):67–79, March 1996.

[42] R. Baeza-Yates and G. Navarro. Block-addressing indices for approximate text retrieval. In *Proc. of the 6th CIKM Conference*, pages 1–8, Las Vegas, Nevada, 1997.

[43] R. Baeza-Yates and G. Navarro. Faster approximate string matching. *Algorithmica*, 23(2):127–158, 1999.

[44] R. Baeza-Yates, G. Navarro, J. Vegas, and P. de la Fuente. A model and a visual query language for structured text. In Berthier Ribeiro-Neto, editor, *Proc. of the 5th Symposium on String Processing and Information Retrieval*, pages 7–13, Santa Cruz, Bolivia, Sept 1998. IEEE CS Press.

[45] R. Baeza-Yates and C. Perleberg. Fast and practical approximate pattern matching. *Information Processing Letters*, 59:21–27, 1996.

[46] Ricardo Baeza-Yates. Visualizing large answers in text databases. In *Int. Workshop on Advanced User Interfaces*, pages 101–107, Gubbio, Italy, May 1996. ACM Press.

[47] Ricardo A. Baeza-Yates, Walter Cunto, Udi Manber, and Sun Wu. Proximity matching using fixed queries trees. In M. Crochemore and D. Gusfield, editors, *5th Combinatorial Pattern Matching*, number 807 in LNCS, pages 198–212. Springer-Verlag, Asilomar, CA, June 1994.

[48] J. Baker. UCLA-NSF Social Aspects of Digital Libraries Workshop, January 1996. http://www.gslis.ucla.edu/DL/.

[49] Y. Bakos and E. Brynjolfsson. Bundling Information Goods: Pricing, Profits, and Efficiency. Technical report, MIT Center for Coordination Science, 1997.

[50] M. Q. W. Baldonado, C.-C. K. Chang, L. Gravano, and A. Paepcke. Metadata for digital libraries: architecture and design rationale. In *Proc. of the 2nd ACM International Conference on Digital Libraries*, pages 47–56, Philadelphia, PA, USA, 1997.

[51] Michelle Q. Wang Baldonado and Terry Winograd. Sensemaker: An information-exploration interface supporting the contextual evolution of a user's interests. In *Proc. of*

the ACM SIGCHI Conference on Human Factors in Computing Systems, pages 11–18, Atlanta, GA, USA, May 1997.

[52] Michelle Q. Wang Baldonado and Terry Winograd. Hi-cites: Dynamically-created citations with active highlighting. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 408–415, Los Angeles, CA, March 1998.

[53] D. Ballard and C. Brown. *Computer Vision.* Prentice Hall, 1982.

[54] Liam Bannon, Allen Cypher, Steven Greenspan, and Melissa L. Monty. Evaluation and analysis of users' activity organization. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 54–57, 1983.

[55] D. Barber. OhioLINK: A Consortial Approach to Digital Library Management. *D-Lib Magazine*, April 1997. http://www.dlib.org/dlib/april97/04barber.html.

[56] E. Barbosa, G. Navarro, R. Baeza-Yates, C. Perleberg, and N. Ziviani. Optimized binary search and text retrieval. In P. Spirakis, editor, *Proc. of European Symposium on Algorithms*, number 979 in LNCS, pages 311–326. Springer-Verlag, 1995. Improved version to appear in *Algorithmica* including W. Cunto as author.

[57] R. Barbosa. *Query Performance on Distributed Digital Libraries.* CS Department, Federal University of Minas Gerais, Brazil, 1998. Master Thesis. In Portuguese.

[58] P. Barford, A. Bestavros, A. Bradley, and M. E. Crovella. Changes in web client access patterns: Characteristics and caching implications. *World Wide Web*, 1999. Special Issue on Characterization and Performance Evaluation, to appear.

[59] P. Barford and M. Crovella. Generating representative Web workloads for network and server performance evaluation. In *ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, pages 151–160, July 1998.

[60] J. Barnett. Computational methods for a mathematical theory of evidence. In *Proc. of the 7th Int. Joint Conference on Artificial Intelligence*, pages 868–875, Vancouver, Canada, 1981.

[61] Jeff Barry, Dania Bilal, and W. David Penniman. Automated system marketplace: the competitive struggle. *Library Journal*, 123(6):43–52, 1998.

[62] Brian T. Bartell, Garrison W. Cottrell, and Richard K. Belew. Latent semantic indexing is an optimal special case of multidimensional scaling. In *Proc. of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 161–167, Copenhagen, Denmark, 1992.

[63] Brian T. Bartell, Garrison W. Cottrell, and Richard K. Belew. Automatic combination of multiple ranked retrieval systems. In *Proc. of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 173–181, Dublin, Ireland, 1994.

[64] Chumki Basu, Haym Hirsh, and William Cohen. Recommendation as classification: Using social and content-based information in recommendation. In *Proc. of AAAI*, pages 714–720, Madison, WI, July 1998.

[65] Marcia J. Bates. The design of browsing and berrypicking techniques for the on-line search interface. *Online Review*, 13(5):407–431, 1989.

[66] Marcia J. Bates. Where should the person stop and the information search interfaces start? *Information Processing & Management*, 26(5):575–591, 1990.

[67] Marcia J. Bates. Document familiarity, relevance, and bradford's law: The Getty online searching project report no. 5. *Information Processing & Management*, 32:697–707, 1996.

[68] Micheline Beaulieu and Christine L. Borgman, editors. *Journal of the American Society for Information Science*, volume 47(7). John Wiley & Sons, 1996.

[69] N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger. The R*-tree: An efficient and robust access method for points and rectangles. *ACM SIGMOD*, pages 322–331, May 1990.

[70] B. Bederson, L. Stead, and J. Hollan. Pad++: Advances in multiscale interfaces. In *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 315–316, Boston, MA, USA, 1994. See this and other papers at http://www.cs.umd.edu/hcil/pad++/papers/ .

[71] Benjamin B. Bederson, James D. Hollan, Ken Perlin, Jonathan Meyer, David Bacon, and George Furnas. Pad++: A zoomable graphical sketchpad for exploring alternate interface physics. *Journal of Visual Languages and Computing*, 7(1):3–31, 1996.

[72] C. Beeri and Y. Kornatzky. A logical query language for hypertext systems. In *Proc. of the European Conference on Hypertext*, pages 67–80. Cambridge University Press, 1990.

[73] Rick Belew. Rave reviews: Acquiring relevance assessments from multiple users. In Marti A. Hearst and Haym Hirsh, editors, *Working Notes of the AAAI Spring Symposium on Machine Learning in Information Access*, Stanford, CA, March 1996.

[74] N. Belkin and B. Croft. Information filtering and information retrieval: Two sides of a same coin? *Communications of the ACM*, 35(12):29–38, December 1992.

[75] N. Belkin, P. G. Marchetti, and C. Cool. Braque – design of an interface to support user interaction in information retrieval. *Information Processing and Management*, 29(3):325–344, 1993.

[76] N. J. Belkin, P. Kantor, E. A. Fox, and J. A. Shaw. Combining the evidence of multiple query representations for information retrieval. *Information Processing & Management*, 31(3):431–448, May–June 1995.

[77] T. C. Bell, A. Moffat, C. Nevill-Manning, I. H. Witten, and J. Zobel. Data compression in full-text retrieval systems. *Journal of the American Society for Information Science*, 44:508–531, 1993.

[78] T. C. Bell, J. Cleary, and I. H. Witten. *Text Compression.* Prentice-Hall, 1990.

[79] Alberto Belussi and Christos Faloutsos. Estimating the selectivity of spatial queries using the 'correlation' fractal dimension. In *Proc. of VLDB Conf.*, pages 299–310, Zurich, Switzerland, September 1995.

[80] A. B. Benítez, M. Beigi, and S-F. Chang. Using relevance feedback in content-based image metasearch. *IEEE Internet Computing*, 2(4):59–69, 1998.

[81] J. Bentley, D. Sleator, R. Tarjan, and V. Wei. A locally adaptive data compression scheme. *Communications of the ACM*, 29:320–330, 1986.

[82] S. Berchtold, C. Boehm, B. Braunmueller, D. A. Keim, and H. P. Kriegel. Fast similarity search in multimedia databases. In *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data*, pages 1–12, 1997.

[83] Stefan Berchtold, Daniel A. Keim, and Hans-Peter Kriegel. The X-tree : An index structure for high-dimensional data. *VLDB*, pages 28–39, 1996.

[84] T. Berners-Lee. The World Wide Web Consortium. http://www.w3.org.

[85] Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. The World-Wide Web. *Communication of the ACM*, 37(8):76–82, 1994.

[86] J. W. Berry. Digital libraries: new initiatives with world wide implications. *IFLA Journal*, 22(1):9–17, 1995. http://www.nlc-bnc.ca/ifla/IV/ifla61/61-berjo.htm.

[87] E. Bertino and E. Ferrari. Temporal synchronization models for multimedia data. *IEEE Trans. on Knowledge and Data Eng.*, 10(4):612–631, July/August 1998.

[88] E. Bertino, E. Ferrari, and M. Stolf. MPGS: An interactive tool for the specification and generation of multimedia presentations. *IEEE Transactions on Multimedia Systems*, 1999. To appear.

[89] E. Bertino and L. D. Martino. *Object-Oriented Database Systems — Concepts and Architecture*. Addison-Wesley, 1993.

[90] E. Bertino and F. Rabitti. Query Processing in MULTOS. In C. Thanos, editor, *Multimedia Office Filing: The MULTOS Approach*, pages 273–295. Elsevier Science Publishers B.V. (North Holland), 1990.

[91] K. Bharat and A. Z. Broder. A technique for measuring the relative size and overlap of public Web search engines. In *7th WWW Conference*, pages 379–388, Brisbane, Australia, 1998.

[92] Krishna Bharat, Andrei Broder, Monika Henzinger, Puneet Kumar, and Suresh Venkatasubramanian. The connectivity server: fast access to linkage information on the Web. In *7th WWW Conf.*, Brisbane, Australia, April 1998.

[93] Krishna Bharat and Monika Henzinger. Improved algorithms for topic distillation in hyperlinked environments. In B. Croft, A. Moffat, C. J. van Rijsbergen, R. Wilkinson, and J. Zobel, editors, *21st ACM SIGIR*, pages 104–111, Melbourne, Australia, 1998.

[94] J. Bichteler and E. A. Eaton III. The combined use of bibliographic coupling and co-citation for document retrieval. *Journal of the American Society for Information Science*, 31(4), July 1980.

[95] Eric A. Bier, Maureen C. Stone, Ken Pier, Ken Fishkin, Thomas Baudel, Matt Conway, William Buxton, and Tony DeRose. Toolglass and magic lenses: The see-through interface. In *Proc. of ACM Conference on Human Factors in Computing Systems*, volume 2 of *Videos: Part II*, pages 445–446, Boston, MA, USA, 1994.

[96] Patricia A. Billingsley. Taking panes: Issues in the design of windowing systems. In Martin Helander, editor, *Handbook of Human-Computer Interaction*, pages 413–436. Springer Verlag, 1988.

[97] W. P. Birmingham, E. H. Durfee, T. Mullen, and M. P. Wellman. The distributed agent architecture of the University of Michigan Digital Library. *AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, 1995. Stanford, CA, AAAI Press.

[98] A. P. Bishop. Working towards an understanding of digital library use: A report on the user research efforts of the NSF/ARPA/NASA DLI projects. *D-Lib Magazine*, October 1995. http://www.dlib.org/dlib/october95/10bishop.html.

[99] K. Bohm and T. C. Rakov. Metadata for multimedia documents. In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, pages 21–26, Minneapolis, MN, USA, May 1994.

[100] Philippe Bonnet and Anthony Tomasic. Partial answers for unavailable data sources. In *Workshop on Flexible Query-Answering Systems*, pages 43–54, 1998.

[101] A. Bookstein. On the perils of merging Boolean and weighted retrieval systems. *Journal of the American Society for Information Sciences*, 29(3):156–158, 1978.

[102] A. Bookstein. Fuzzy requests: An approach to weighted Boolean searches. *Journal of the American Society for Information Sciences*, 31:240–247, 1980.

[103] A. Bookstein. Implication of Boolean structure for probabilistic retrieval. In *Proc. of the 8th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pages 11–17, Montreal, Canada, 1985.

[104] Christine L. Borgman. Why are online catalogs hard to use? Lessons learned from information retrieval studies. *Journal of the American Society for Information Science*, 37(6):387–400, 1986.

[105] Christine L. Borgman. Why are online catalogs *still* hard to use? *Journal of the American Society for Information Science*, 47(7):493–503, 1996.

[106] J. Bosak. XML, Java, and the future of the Web. Technical report, Sun Microsystems, 1997. http://sunsite.unc.edu/pub/sun-info/standards/xml/why/xmlapps.htm.

[107] Charles P. Bourne. On-line systems: history, technology, and economics. *Journal of the American Society for Information Science*, 31:155–160, 1980.

[108] C. Mic Bowman, Peter B. Danzig, Darren R. Hardy, Udi Manber, and Michael F. Schwartz. The Harvest information discovery and access system. In *Proc. 2nd Int. WWW Conf.*, pages 763–771, October 1994.

[109] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis: Forecasting and Control.* Prentice Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1994.

[110] R. S. Boyer and J. S. Moore. A fast string searching algorithm. *Communications of the ACM*, 20(10):762–772, 1977.

[111] James Boyle, William Ogden, Steven Uhlir, and Patricia Wilson. QMF usability: How it really happened. In *Proc. of IFIP INTERACT'84: Human-Computer Interaction*, pages 877–882, 1984.

[112] Tolga Bozkaya and Z. Meral Ozsoyoglu. Distance-based indexing for high-dimensional metric spaces. In *Proc. of ACM SIGMOD Conference*, pages 357–368, Tucson, AZ, USA, 1997.

[113] P. M. E. De Bra and R. D. J. Post. Searching for arbitrary information in the WWW: The fish search for Mosaic. In *Proc. of the 2nd Int. WWW Conference*, Chicago, October 1994. http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/www-fall94.html.

[114] T. Bray. Measuring the web. In *Proc. of the 5th Int. WWW Conference*, Paris, May 1996. http://www5conf.inria.fr/fich_html/papers/P9/Overview.html.

[115] S. Brin. Extracting patterns and relations from the World Wide Web. In *Workshop on Web Databases*, Valencia, Spain, March 1998.

[116] Sergei Brin. Near neighbor search in large metric spaces. In *Proc. of VLDB Conf.*, pages 574–584, Zurich, Switzerland, Sept 1995.

[117] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. In *Proc. of the 7th Int. WWW Conference*, Brisbane, Australia, April 1998.

[118] T. Brinkhoff, H. P. Kriegel, and R. Schneider. Comparison of approximations of complex objects used for approximation-based query processing in spatial database systems. In *Proc. of the IEEE Conference on Data Engineering*, pages 40–49, Vienna, Austria, 1993.

[119] Thomas Brinkhoff, Hans-Peter Kriegel, Ralf Schneider, and Bernhard Seeger. Multi-step processing of spatial joins. In *Proc. of ACM SIGMOD*, pages 197–208, Minneapolis, MN, USA, May 1994.

[120] A. Broder, S. Glassman, M. Manasse, and G. Zweig. Syntactic clustering of the Web. In *6th Int. WWW Conference*, pages 391–404, Santa Clara, CA, USA, April 1997.

[121] A. Z. Broder. On the resemblance and containment of documents. In *Conf. on Compression and Complexity of Sequences*, pages 21–29, Salerno, Italy, 1997. IEEE Computer Society.

[122] J. Broglio, J. P. Callan, W. B. Croft, and D. W. Nachbar. Document retrieval and routing using the INQUERY system. In D. K. Harman, editor, *Overview of the Third Retrieval Conference (TREC-3)*, pages 29–38. NIST Special Publication 500-225, 1995.

[123] Eric W. Brown. *Execution Performance Issues in Full-Text Information Retrieval.* PhD thesis, University of Massachusetts, Amherst, 1996. Available as UMass Comp. Sci. Tech. Rep. TR95-81.

[124] J. S. Brown and P. Duguid. The social life of documents. *First Monday*, May 1996. http://www.firstmonday.dk/issues/issue1/documents/.

[125] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, December 1992.

[126] M. Buckland and C. Plaunt. On the Construction of Selection Systems. *Library Hi Tech*, 12:15–28, 1994.

[127] Chris Buckley, Gerard Salton, and James Allan. The effect of adding relevance information in a relevance feedback environment. In *Proc. of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 292–300, Dublin, Ireland, 1994.

[128] P. Buneman, S. B. Davidson, K. Hart, C. Overton, and L. Wong. A data transformation system for biological data sources. In *Proc. of the 21st International Conference on Very Large Data Bases*, pages 158–169, Zurich, Switzerland, September 1995.

[129] Peter Buneman. Semistructured data. In *ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 117–121, Tucson, Arizona, 1997.

[130] Robin Burke, Kristian Hammond, Vladimir Kulukin, Steven Lytinen, Noriko Tomuro, and Scott Schoenberg. Experiences with the FAQ system. *AI Magazine*, 18(2):57–66, 1997.

[131] W. A. Burkhard and R. M. Keller. Some approaches to best-match file searching. *Communications of the ACM*, 16(4):230–236, April 1973.

[132] F. Burkowski. An algebra for hierarchically organized text-dominated databases. *Information Processing & Management*, 28(3):333–348, 1992.

[133] F. Burkowski. Retrieval activities in a database consisting of heterogeneous collections of structured text. In *Proc. of the 15th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pages 112–125, Copenhagen, Denmark, 1992.

[134] Arthur W. Burks, Herman H. Goldstine, and John von Neumann. Preliminary discussion of the logical design of an electronic computing instrument. In William Aspray and Arthur Burks, editors, *Papers of John von Neumann on Computers and Computer Theory*, pages 97–142. The MIT Press, Cambridge, MA, 1987 (originally appeared in 1946).

[135] V. Bush. As we may think. *The Atlantic Monthly*, 176(1):101–108, July 1945.

[136] D. Byers. Full-text indexing of non-textual resources. In *7th WWW Conf.*, Brisbane, Australia, April 1998.

[137] Brendon Cahoon and Kathryn McKinley. Performance evaluation of a distributed architecture for information retrieval. In *Proc. 19th Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 110–118, Zurich, Switzerland, August 1996.

[138] J. Callan. Document filtering with inference networks. In *Proc. of the 19th ACM SIGIR Conference*, pages 262–269, Zurich, Switzerland, August 1996.

[139] J. P. Callan, Z. Lu, and W. Bruce Croft. Searching Distributed Collections with Inference Networks. In E. A. Fox, P. Ingwersen, and R. Fidel, editors, *Proc. of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 21–28, Seattle, Washington, July 1995. ACM Press.

[140] S. K. Card, M. Pavel, and J. E. Farrell. Window-based computer dialogues. In *Proc. of Interact '84, First IFIP Conference on Human-Computer Interaction*, pages 355–359, 1984.

[141] Stuart K. Card, Jock D. MacKinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision to Think.* Morgan Kaufmann Publishers, 1998.

[142] Stuart K. Card, Thomas P. Moran, and Allen Newell. *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates, 1983.

[143] Stuart K. Card, George G. Robertson, and Jock D. Mackinlay. The information visualizer, an information workspace. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 181–188, 1991.

[144] Stuart K. Card, George G. Robertson, and William York. The WebBook and the Web Forager: An information workspace for the World-Wide Web. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 111–117, Zurich, Switzerland, April 1996.

[145] S. Carliner. Designing wizards. *Training & Development*, 52(7):62–63, 1998.

[146] Carnegie Mellon University DLI Team, Informedia. http://www.informedia.cs.cmu.edu/.

[147] Claudio Carpineto and Giovanni Romano. Information retrieval through hybrid navigation of lattice representations. *International Journal of Human-Computer Studies*, 45(5):553–578, 1996.

[148] J. Carriere and R. Kazman. WebQuery: Searching and visualizing the Web through connectivity. In *6th Int. WWW Conf.*, Santa Clara, CA, USA, April 1997.

[149] M. Castagli and S. Eubank. *Nonlinear Modeling and Forecasting*. Addison-Wesley, 1992.

[150] R. Cattel. *The Object Database Standard: ODMG-93*. Morgan Kauffman, 1993.

[151] W. B. Cavnar and A. M. Gillies. Data retrieval and the realities of document donversion. In *Proc. of the 1st Annual Conference on the Theory and Practice of Digital Libraries*, College Station, TX, USA, June 1994.

[152] S. Chakrabarti, B. Dom, R. Agrawal, and P. Raghavan. Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies. *VLDB Journal*, 7(3):163–178, 1998.

[153] S. Chakrabarti, B. Dom, D. Gibson, S. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Experiments in topic distillation. In *ACM-SIGIR'98 Post Conference Workshop on Hypertext Information Retrieval for the Web*, Melbourne, Australia, 1998.

[154] S. Chakrabarti, B. Dom, P. Raghavan, S. Rajagopalan, D. Gibson, and J. Kleinberg. Automatic resource compilation by analyzing hyperlink structure and associated text. In *7th Int. WWW Conference*, pages 65–74, Brisbane, Australia, April 1998.

[155] M. Chalmers, K. Rodden, and D. Brodbeck. The order of things: activity-centred information access. In *7th Int. WWW Conference*, Brisbane, Australia, April 1998.

[156] Matthew Chalmers and Paul Chitson. Bead: Exploration in information visualization. In *Proc. of the 15th Annual International ACM/SIGIR Conference*, pages 330–337, Copenhagen, Denmark, 1992.

[157] C-H. Chang and C-C. Hsu. Customizable multi-engine search tool with clustering. In *6th. Int. WWW Conference*, Santa Clara, CA, USA, April 1997.

[158] S.-F. Chang, J. R. Smith, H. J. Meng, H. Wang, and D. Zhong. Finding Images/Video in Large Archives: Columbia's Content-Based Visual Query Project. *D-Lib Magazine*, February 1997. http://www.dlib.org/dlib/february97/columbia/02chang.html.

[159] Shan-Ju Chang and Ronald E. Rice. Browsing: A multidimensional framework. *Annual Review of Information Science and Technology*, 28:231–276, 1993.

[160] S. Chapman and A. R. Kenney. Digital conversion of library research materials: A case for full informational capture. *D-Lib Magazine*, October 1996. http://www.dlib.org/dlib/october96/cornell/10chapman.html.

[161] S. Chawathe, H. García-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman, and J. Widom. The TSIMMIS project: Integration of heterogeneous information sources. In *Proc. of IPSJ Conference*, pages 7–18, October 1994.

[162] C. Chen. Structuring and visualizing the WWW by generalized similarity analysis. In *Proc. of the 8th ACM Conference on Hypertext and Hypermedia*, pages 177–186, Southampton, England, 1997.

[163] Hsinchun Chen, Andrea L. Houston, Robin R. Sewell, and Bruce R. Schatz. Internet browsing and searching: User evaluations of category map and concept space techniques. *Journal of the American Society for Information Science*, 49(7):582–608, 1998.

[164] Michael Chen and Marti A. Hearst. Presenting Web site search results in context: A demonstration. In *Proc. of the 20th Annual International ACM/SIGIR Conference*, page 381, Melbourne, Australia, 1998.

[165] Peter M. Chen, Edward K. Lee, Garth A. Gibson. Randy H. Katz, and David A. Paterson. RAID: High-performance, reliable secondary storage. *ACM Computing Surveys*, 26(2):145–185, June 1994.

[166] F. Cheong. *Internet Agents, Spiders, Wanderers, Brokers and Bots.* New Riders, 1996.

[167] D. R. Chestnutt. The model editions partnership: 'Smart Text' and beyond. *D-Lib Magazine*, July/August 1997. http://www.dlib.org/dlib/july97/07chesnutt.html.

[168] J. Cho, H. García-Molina, and L. Page. Efficient crawling through URL ordering. In *7th WWW Conference*, Brisbane, Australia, April 1998.

[169] S. Christodoulakis, M. Theodoridou, F. Ho, M. Papa, and A. Pathria. Multimedia document presentation, information extraction, and document formation in MINOS: a model and a system. *IEEE Transactions on Office Information Systems*, 4(4):345–383, October 1986.

[170] W. W. Chu, I. T. Ieong, R. K. Taira, and C. M. Breant. A Temporal Evolutionary Object-Oriented Model and its Query Languages for Medical Image Management. In *Proc. of the Int. Conference on Very Large Data Bases*, pages 53–64, Vancouver, Canada, Aug 1992.

[171] K. Church and W. Gale. Poisson mixtures. *Natural Language Engineering*, 1(2):163–190, 1995.

[172] Paolo Ciaccia, Marco Patella, and Pavel Zezula. M-tree: An efficient access method for similarity search in metric spaces. In *Proc. of VLDB Conf.*, pages 426–435, Athens, Greece, August 1997.

[173] C. Clarke, G. Cormack, and F. Burkowski. An algebra for structured text search and a framework for its implementation. *The Computer Journal*, 38(1):43–56, 1995.

[174] Charles L. A. Clarke, Gordon V. Cormack, and Forbes J. Burkowski. Shortest substring ranking (multitext experiments for TREC-4). In Donna Harman, editor, *Proc. of the Fourth Text Retrieval Conference TREC-4*. National Institute of Standards and Technology Special Publication, 1996.

[175] E. G. Coffman, Z. Liu, and R. R. Weber. Optimal robot scheduling for Web search engines. Technical Report 3317, INRIA, France, December 1997.

[176] Douglas E. Comer and David L. Stevens. *Internetworking with TCP/IP Vol III: Client-Server Programming and Applications.* Prentice-Hall, Inc., Englewood Cliffs, NJ, USA, 1993.

[177] *Communications of the ACM*, 37(2), February 1994.

[178] *Communications of the ACM*, 38(8), August 1995.

[179] B. Commentz-Walter. A string matching algorithm fast on the average. In *Proc. ICALP'79*, pages 118–132. Springer-Verlag, 1979.

[180] Compaq: Junglee shopping guide. http://www.compaq.junglee/compaq/top.html, 1998.

[181] Jeff Conklin. Hypertext: An introduction and survey. *IEEE Computer*, 20(9):17–41, September 1987.

[182] Dan Connolly on the architecture of the Web: Let a thousand flowers bloom. *IEEE Internet Computing*, 2(2):22–31, 1998.

[183] M. Consens and T. Milo. Algebras for querying text regions. In *Proc. of the ACM Conference on Principles of Distributed Systems*, pages 11–22, San Jose, California, USA, 1995.

[184] M. P. Consens and A. O. Mendelzon. Expressing structural hypertext queries in Graphlog. In *Hypertext'89*, pages 269–292, 1989.

[185] Alan Cooper. *About Face: The Essentials of User Interface Design.* IDG Books, 1995.

[186] William S. Cooper, Fredric C. Gey, and Daniel P. Dabney. Probabilistic retrieval based on staged logistic regression. In *Proc. of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 198–210, Copenhagen, Denmark, 1992.

[187] Wm. S. Cooper. The formalism of probability theory in IR: A foundation or an encumbrance? In *Proc. of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 242-247, Dublin, Ireland, 1994. Triennial ACM-SIGIR Award Paper.

[188] Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms.* The MIT Press/McGraw-Hill, Cambridge, MA, 1990.

[189] Corporation for National Research Initiatives, The Handle System, May 1998. http://www.handle.net/.

[190] P. Cotton, editor. *ISO Working Draft SQL Multimedia and Application Packages (SQL/MM), Part II Full text.* ISO/IEC, 1993. SC21/WG3 N1631, SQL/MM MUN-004.

[191] S. B. Cousins, S. P. Ketchpel, A. Paepcke, H. García-Molina, S. W. Hassan, and M. Roscheisen. InterPay: Managing multiple payment mechanisms in digital libraries. In *Proc. of the 2nd Annual Conference on the Theory and Practice of Digital Libraries*, Austin, Texas, June 1995.

[192] S. B. Cousins, A. Paepcke, T. Winograd, E. A. Bier, and K. Pier. The digital library integrated task environment (DLITE). In *Proc. of the 2nd ACM International Conference on Digital Libraries*, pages 142–151, Philadelphia, PA, USA, July 1997.

[193] Steve B. Cousins. *Reification and Affordances in a User Interface for Interacting with Heterogeneous Distributed Applications.* PhD thesis, Stanford University, May 1997.

[194] B. Cox, J. D. Tygar, and M. Sirbu. NetBill Security and Transaction Protocol. In *Proc. of the 1st USENIX Workshop on Electronic Commerce*, New York, USA, July 1995.

[195] S. Y. Crawford, J. M. Hurd, and A. C. Weller. *From Print to Electronic: the Transformation of Scientific Communication.* American Society for Information Science Monograph Series, Medford, New Jersey, 1996.

[196] M. Crochemore and W. Rytter. *Text Algorithms.* Oxford University Press, Oxford, UK, 1994.

[197] W. B. Croft. What do people want from information retrieval? (The top 10 research issues for companies that use and sell IR systems). *D-Lib Magazine*, November 1995. http://www.dlib.org/dlib/november95/11croft.html.

[198] W. B. Croft. Experiments with representation in a document retrieval system. *Information Technology: Research and Development*, 2(1):1–21, 1983.

[199] W. B. Croft and D. J. Harper. Using probabilistic models of retrieval without relevance information. *Journal of Documentation*, 35(4):285–295, 1979.

[200] Carolyn J. Crouch and Bokyung Yang. Experiments in automatic statistical thesaurus construction. In *Proc. of the ACM-SIGIR Conference on Research and Development in Information Retrieval*, pages 77–88, Copenhagen, Denmark, 1992.

[201] M. Crovella and A. Bestavros. Self-similarity in World Wide Web traffic: Evidence and possible causes. In *ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, pages 160–169, May 1996.

[202] Douglass R. Cutting, David Karger, and Jan Pedersen. Constant interaction-time Scatter/Gather browsing of very large document collections. In *Proc. of the 16th Annual International ACM/SIGIR Conference*, pages 126–135, Pittsburgh, PA, 1993.

[203] Douglass R. Cutting, Jan O. Pedersen, David Karger, and John W. Tukey. Scatter/Gather: A cluster-based approach to browsing large document collections. In *Proc. of the 15th Annual International ACM/SIGIR Conference*, pages 318–329, Copenhagen, Denmark, 1992.

[204] Allen Cypher, editor. *Watch What I Do: Programming by Demonstration*. MIT Press, Cambridge, 1993.

[205] A. Czumaj, M. Crochemore, L. Gasieniec, S. Jarominek, Thierry Lecroq, W. Plandowski, and W. Rytter. Speeding up two string-matching algorithms. *Algorithmica*, 12:247–267, 1994.

[206] T. Dao, R. Sacks-Davis, and J. Thom. Indexing structured text for queries on containment relationships. In *Proc. of 7th Australasian Database Conference*, Melbourne, Australia, Jan/Feb 1996.

[207] Rick Darnell. *HTML 4.0 Unleashed, Professional Reference Edition*. Samms.net Publishing, 1998.

[208] M. Dartois, A. Maeda, T. Sakaguchi, T. Fujita, S. Sugimoto, and K. Tabata. A multilingual electronic text collection of folk tales for casual users using off-the-shelf browsers. *D-Lib Magazine*, October 1997. http://www.dlib.org/dlib/october97/sugimoto/10sugimoto.html.

[209] S. Davidson, C. Overton, V. Tannen, and L. Wong. BioKleisli: A digital library for biomedical researchers. *International Journal of Digital Libraries*, 1(1):35–53, April 1997.

[210] J. R. Davis and C. Lagoze. A protocol and server for a distributed technical report library. Technical report, Cornell University Computer Science Department, June 1994.

[211] Y. F. Day, S. Dagtas, M. Iino, A. Khokhar, and A. Ghafoor. Object-oriented conceptual modeling of video data. In *Proc. of the IEEE Conference on Data Engineering*, pages 401–405, Taipei, Taiwan, 1995.

[212] DejaNews: Main Page. http://www.dejanews.com, 1998.

[213] Alin Deutsch, Mary Fernández, Daniela Florescu, Alon Levy, and Dan Suciu. A query language for XML. http://www.research.att.com/~mff/xml/w3cnote.html, 1998.

[214] A. Dillon. What is the shape of information? Human factors in the development and use of digital libraries. Allerton discussion document submitted for the 1995 Allerton Institute. http://edfu.lis.uiuc.edu/allerton/95/s4/dillon.html.

[215] DirectHit: Main Page. http://www.directhit.com, 1998.

[216] Division of Information and Intelligent Systems, National Science Foundation. Digital Libraries Initiative – Phase 2. Announcement Number NSF 98-63, 1998. http://-www.nsf.gov/pubs/1998/nsf9863/nsf9863.htm.

[217] M. Dodge. The geography of cyberspace directory: Main page. http:-//www.geog.ucl.ac.uk/casa/martin/geography_of_cyberspace.html, 1997.

[218] R. Dolin, D. Agrawal, A. El Abbadi, and J. Pearlman. Using automated classification for summarizing and selecting heterogeneous information sources. *D-Lib Magazine*, January 1998. http://www.dlib.org/dlib/january98/dolin/01dolin.html.

[219] R. Dolin, D. Agrawal, L. Dillon, and A. El Abbadi. Pharos: a scalable distributed architecture for locating heterogeneous information sources. In *Proc. of the 6th CIKM Conference*, pages 348–355, Las Vegas, Nevada, 1997.

[220] P. Dömel. Webmap: A graphical hypertext navigation tool. In *Proc. of the 2nd. WWW Conference*, 1994. http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/-WWW2_Proceedings.html.

[221] P. Doty and A. P. Bishop. The national information infrastructure and electronic publishing: A reflective essay. *Journal of the American Society for Information Science*, 45(10):785–799, 1994.

[222] Karen M. Drabenstott and Marjorie S. Weller. The exact-display approach for online catalog subject searching. *Information Processing and Management*, 32(6):719–745, 1996.

[223] D. Dreilinger. Savvy Search, 1996. http://savvy.cs.colostate.edu:2000/form?beta.

[224] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.

[225] Mark D. Dunlop. The effect of accessing nonmatching documents on relevance feedback. *ACM Transactions on Information Systems*, 15(2):137–153, 1997.

[226] Deborah M. Edwards and Lynda Hardman. 'Lost in hyperspace': Cognitive mapping and navigation in a hypertext environment. In Ray McAleese, editor, *Hypertext I: Theory into Practice*, pages 105–125. Ablex Publishing Corporation, 1988.

[227] Dennis E. Egan. Individual differences in human-computer interaction. In Martin Helander, editor, *Handbook of Human-Computer Interaction*, pages 543–568. Springer Verlag, 1988.

[228] Dennis E. Egan, Michael E. Lesk, R. Daniel Ketchum, Carol C. Lochbaum, Joel R. Remde, Louis M. Gómez, and Thomas K. Landauer. Hypertext for the electronic library? CORE sample results. In *Proc. of the ACM Hypertext Conference*, pages 299–312, May 1991.

[229] Dennis E. Egan, Joel R. Remde, Louis M. Gomez, Thomas K. Landauer, Jennifer Eberhardt, and Carol C. Lochbaum. Formative design evaluation of SuperBook. *Transactions on Information Systems*, 7(1):30–57, 1989.

[230] Dennis E. Egan, Joel R. Remde, Thomas K. Landauer, Carol C. Lochbaum, and Louis M. Gomez. Behavioral evaluation and analysis of a hypertext browser. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 205–210, May 1989.

[231] M. J. Egenhofer. Extending SQL for Graphical Display. *Cartography and Geographic Information Systems*, 18(4):230–245, 1991.

[232] Stephen G. Eick. Graphically displaying text. *Journal of Computational and Graphical Statistics*, 3(2):127–142, June 1994.

[233] Stephen G. Eick and Graham J. Wills. High interaction graphics. *European Journal of Operations Research*, 81(3):445–459, March 1995.

[234] Stephen G. Eick and Graham J. Wills. Navigating large networks with hierarchies. In *Proc. of the Conference on Visualization '93*, pages 204–209, San Jose, October 1993.

[235] P. Elias. Universal codeword sets and representations of the integers. *IEEE Transactions on Information Theory*, 21:194–203, 1975.

[236] D. Ellis. A behavioural model for information retrieval system design. *Journal of Information Science*, 15:237–247, 1989.

[237] R. Entlich, L. Garson, M. Lesk, L. Normore, J. Olsen, and S. Weibel. Making a digital library: The Chemistry Online Retrieval Experiment – A summary of the CORE project (1991-1995). *D-Lib Magazine*, December 1995. http://www.dlib.org/dlib/december95/briefings/12core.html.

[238] M. L. Escobar-Molano and S. Ghandeharizadeh. A Framework for Conceptualizing Structured Video. In *Proc. First Int. Workshop on Multimedia Information Systems*, pages 95–110, Arlington,Virginia, September 1995.

[239] ERCIM Digital Library Working Group, European Research Consortium for Informatics and Mathematics, June 1998. http://www.area.pi.cnr.it/ErcimDL/.

[240] Excite: Main page. http://www.excite.com, 1995.

[241] Express: Main Page. http://www.express.infoseek.com, 1998.

[242] C. Faloutsos and R. Chan. Text access methods for optical and large magnetic disks: design and performance comparison. In *Proc. of VLDB'88*, pages 280–293, Los Angeles, CA, USA, 1988.

[243] C. Faloutsos and S. Christodoulakis. Description and performance analysis of signature file methods. *ACM TOIS*, 5(3):237–257, 1987.

[244] Christos Faloutsos, Ron Barber, Myron Flickner, J. Hafner, Wayne Niblack, Dragutin Petkovic, and William Equitz. Efficient and effective querying by image content. *J. of Intelligent Information Systems*, 3(3/4):231–262, July 1994.

[245] Christos Faloutsos and Volker Gaede. Analysis of n-dimensional quadtrees using the Hausdorff fractal dimension. In *Proc. of VLDB Conf.*, pages 40–50, Bombay, India, September 1996.

[246] Christos Faloutsos, H. V. Jagadish, Alberto O. Mendelzon, and Tova Milo. A signature technique for similarity-based queries. In *SEQUENCES'97*, Salerno, Italy, June 1997.

[247] Christos Faloutsos and Ibrahim Kamel. Beyond uniformity and independence: Analysis of R-trees using the concept of fractal dimension. In *Proc. ACM SIGACT-SIGMOD-SIGART PODS*, pages 4–13, Minneapolis, MN, May 1994.

[248] Christos Faloutsos and King-Ip (David) Lin. FastMap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. *Proc. of ACM SIGMOD*, pages 163–174, May 1995.

[249] Christos Faloutsos, M. Ranganathan, and Yannis Manolopoulos. Fast subsequence matching in time-series databases. In *Proc. ACM SIGMOD*, pages 419–429, Minneapolis, MN, USA, May 1994. 'Best Paper' award.

[250] M. Farach and M. Thorup. String matching in Lempel-Ziv compressed strings. In *Proc. 27th ACM Annual Symposium on the Theory of Computing*, pages 703–712, Las Vegas, NE, USA, 1995.

[251] I. A. Ferguson and M. J. Wooldridge. Paying their way: Commercial digital libraries for the 21st century. *D-Lib Magazine*, June 1997. http://www.dlib.org/dlib/june97/zuno/06ferguson.html.

[252] Terry Ellen Ferl and Larry Millsap. The knuckle-cracker's dilemma: a transaction log study of OPAC subject searching. *Information Technology and Libraries*, 15:81–98, 1996.

[253] Mary Fernandez, Daniela Florescu, Alon Levy, and Dan Suciu. A query language for a Web-site management system. *SIGMOD Record*, 26(3):4–11, September 1997.

[254] E. Ferrari and G. Haus. Designing music objects for a multimedia database. In *Proc. 12th Colloquium on Musical Informatics*, pages 135–138, Gorizia, Italy, September 1998.

[255] Gerhard Fischer and Helga Nieper-Lemke. Helgon: Extending the retrieval reformulation paradigm. In *Proc. of ACM CHI Conference on Human Factors in Computing Systems*, pages 357–362, 1989.

[256] Ken Fishkin and Maureen C. Stone. Enhanced dynamic queries via movable filters. In *Proc. of ACM CHI Conference on Human Factors in Computing Systems*, volume 1 of *Papers: Information Visualization*, pages 415–420, Denver, CO, USA, 1995.

[257] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jon Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 28(9):23–32, September 1995.

[258] D. Florescu, A. Levy, and A. Mendelzon. Database techniques for the World-Wide Web: A survey. *SIGMOD Record*, 27(3):59–74, 1998.

[259] M. J. Flynn. Very high-speed computing systems. In *Proc. IEEE*, volume 54, pages 1901–1909, 1966.

[260] Corporation for National Research Initiatives. Computer Science Technical Reports Project (CSTR), May 1996. http://www.cnri.reston.va.us/home/cstr.html.

[261] D. J. Foskett. Thesaurus. In K. Sparck Jones and P. Willet, editors, *Readings in Information Retrieval*, pages 111–134. Morgan Kaufmann Publishers, Inc., 1997.

[262] Richard H. Fowler, Wendy A. L. Fowler, and Bradley A. Wilson. Integrating query, thesaurus, and documents through a common visual representation. In *Proc. of the 14th Annual International ACM/SIGIR Conference*, pages 142–151, Chicago, 1991.

[263] C. Fox. Lexical analysis and stoplists. In W. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures & Algorithms*, pages 102–130. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[264] E. Fox, R. France, E. Sahle, A. Daoud, and B. Cline. Development of a Modern OPAC: From REVTOLC to MARIAN. In *Proc. of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 248–259, Pittsburgh, PA, June/July 1993.

[265] E. Fox and L. Lunin. Introduction and overview to perspectives on digital libraries: Guest editor's introduction to special issue. *Journal of the American Society for Information Science*, 44(8):441–443, 1993.

[266] E. A. Fox. Source book on digital libraries. Technical Report TR-93-35, Virginia Polytechnic Institute and State University, 1993.

[267] E. A. Fox, R. M. Akscyn, R. K. Furuta, and J. J. Leggett. Digital libraries. *Communications of the ACM*, 38(4):22–28, April 1995.

[268] E. A. Fox and R. Gupta. Courseware on Digital Libraries. http://ei.cs.vt.edu/~dlib/.

[269] E. A. Fox, L. S. Heath, and D. Hix. Project Envision Final Report: A user-centered database from the computer science literature, July 1995. http://ei.cs.vt.edu/papers/ENVreport/final.html.

[270] E. A. Fox and G. Marchionini. Toward a worldwide digital library. *Communications of the ACM*, 41(4):29–32, April 1998. http://purl.lib.vt.edu/dlib/pubs/CACM199804.

[271] E. A. Fox. *Extending the Boolean and Vector Space Models of Information Retrieval with P-Norm Queries and Multiple Concept Types*. PhD thesis, Cornell University, Ithaca, New York, Http://www.ncstrl.org, 1983.

[272] Edward A. Fox. Characterization of two new experimental collections in computer and information science containing textual and bibliographical concepts. *Technical Report*, 83-561, 1983. Http://www.ncstrl.org.

[273] Edward A. Fox, Deborah Hix, Lucy T. Nowell, Dennis J. Brueni, William C. Wake, Lenwood S. Heath, and Durgesh Rao. Users, user interfaces, and objects: Envision, a digital library. *Journal of the American Society for Information Science*, 44(8):480–491, 1993.

[274] W. Frakes. Stemming algorithms. In W. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures & Algorithms*, pages 131–160. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[275] W. B. Frakes and R. Baeza-Yates. *Information Retrieval: Data Structures & Algorithms.* Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[276] W. Francis and H. Kucera. *Frequency Analysis of English Usage.* Houghton Mifflin Co., 1982.

[277] Eric Freeman and Scott Fertig. Lifestreams: Organizing your electronic life. In Robin Burke, editor, *Working Notes of the AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, Cambridge, MA, November 1995.

[278] H. P. Frei and D. Stieger. The use of semantic links in hypertext information retrieval. *Information Processing & Management*, 31(1):1–13, 1994.

[279] J. French, E. Fox, K. Maly, and A. Selman. Wide Area Technical Report Service: Technical Reports Online. *Communications of the ACM*, 38(4):47, April 1995.

[280] A. Friedlander. D-lib Program: Research in Digital Libraries, May 1998. http://www.dlib.org/.

[281] N. Fuhr. Models for retrieval with probabilistic indexing. *Information Processing & Management*, 25:55–72, 1989.

[282] N. Fuhr. Probabilistic models in information retrieval. *The Computer Journal*, 35(3):243–255, 1992.

[283] N. Fuhr. DOLORES: A System for Logic-Based Retrieval of Multimedia Objects. In B. Croft, A. Moffat, C. J. van Rijsbergen, R. Wilkinson, and J. Zobel, editors, *Proc. of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 257–265, Melbourne, Australia, 1998.

[284] Norbert Fuhr. Optimal polynomial retrieval functions based on the probability ranking principle. *ACM Transactions on Information Systems*, 7(3):183–204, 1989.

[285] Keinosuke Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, 2nd edition, 1990.

[286] George W. Furnas and Jeff Zacks. Multitrees: Enriching and reusing hierarchical structure. In *Proc. of ACM CHI Conference on Human Factors in Computing Systems*, volume 2, pages 330–336, 1994.

[287] G. W. Furnas, S. Deerwester, S. T. Dumais, T. K. Landauer, R. A. Harshman, L. A. Streeter, and K. E. Lochbaum. Information retrieval using a singular value decomposition model of latent semantic structure. In *Proc. of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 465–480, 1988.

[288] R. Furuta. Defining and Using Structure in Digital Documents. In *Proc. of the 1st Annual Conference on the Theory and Practice of Digital Libraries*, College Station, TX, USA, 1994.

[289] Richard Furuta, Frank M. Shipman III, Catherine C. Marshall, Donald Brenner, and Hao wei Hsieh. Hypertext paths and the World-Wide Web: Experiences with Walden's paths. In *Proc. of the Eighth ACM Conference on Hypertext*, pages 167–176, Southampton, England, 1997.

[290] Volker Gaede and Oliver Günther. Multidimensional access methods. *ACM Computing Surveys*, 30(2):170–231, 1998.

[291] R. G. Gallager. Variations on a theme by Huffman. *IEEE Transactions on Information Theory*, 24:668–674, 1978.

[292] S. Gauch, G. Wang, and M. Gómez. ProFusion: Intelligent fusion from multiple, distributed search engines. *Journal of Universal Computing*, 2(9):637–649, September 1996.

[293] E. Gaynor. From MARC to Markup: SGML and Online Library Systems. *ALCTS Newsletter*, 7(2), 1996.

[294] Naum Gershon, Joshua LeVasseur, Joel Winstead, James Croall, Ari Pernick, and William Ruh. Visualizing Internet resources. In *Proc. of the '95 Information Visualization*, pages 122–128, Atlanta, October 1995.

[295] Fredric C. Gey. Inferring probability of relevance using the method of logistic regression. In *Proc. of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 222–231, Dublin, Ireland, 1994.

[296] S. Gibbs. Composite Multimedia and Active Objects. In *Proc. Int. Conf. on Object-Oriented Programming: Systems, Languages, and Applications*, pages 97–112, October 1991.

[297] S. Gibbs, C. Breiteneder, and D. Tsichritzis. Audio/video databases: an object-oriented approach. In *Proc. 9th Int. Conf. on Data Engineering*, pages 381–390, 1993.

[298] D. Gibson, J. Kleinberg, and P. Raghavan. Inferring Web communities from link topologies. In *9th ACM Conference on Hypertext and Hypermedia*, Pittsburgh, USA, 1998.

[299] H. Gladney, E. Fox, Z. Ahmed, R. Ashany, N. Belkin, and M. Zemankova. Digital library: Gross structure and requirements: Report from a March 1994 workshop. In *Proc. of the 1st Annual Conference on the Theory and Practice of Digital Libraries*, pages 101–107, College Station, TX, USA, 1994.

[300] H. Gladney, F. Mintzer, F. Schiattarella, J. Bescós, and M. Treu. Digital access to antiquities. *Communications of the ACM*, 41(4):49–57, April 1998.

[301] H. M. Gladney. Safeguarding digital library contents and users: document access control. *D-Lib Magazine*, June 1997. http://www.dlib.org/dlib/june97/ibm/06gladney.html.

[302] H. M. Gladney and J. B. Lotspiech. Safeguarding digital library contents and users: Assuring convenient security and data quality. *D-Lib Magazine*, May 1997. http://www.dlib.org/dlib/may97/ibm/05gladney.html.

[303] Charles Goldfarb. *The SGML Handbook*. Oxford University Press, Oxford, 1990.

[304] Charles Goldfarb and Paul Prescod. *The XML Handbook*. Prentice Hall, Oxford, 1998.

[305] Dina Q. Goldin and Paris C. Kanellakis. On similarity queries for time-series data: Constraint specification and implementation. *Int. Conf. on Principles and Practice of Constraint Programming*, pages 137–153, Sept 1995.

[306] R. Goldman, J. McHugh, and J. Widom. Lore: A database management system for XML. Technical report, Stanford University Database Group, 1998.

[307] S. W. Golomb. Run-length encodings. *IEEE Transactions on Information Theory*, 12(3):399–401, 1966.

[308] G. Golovchinsky. What the query told the link: the integration of Hypertext and Information Retrieval. In *8th ACM Conference on Hypertext and Hypermedia*, pages 67–74, Southampton, England, 1997.

[309] G. Gonnet. Examples of PAT applied to the Oxford English Dictionary. Technical Report OED-87-02, UW Centre for the New OED and Text Research, Univ. of Waterloo, 1987.

[310] G. Gonnet and R. Baeza-Yates. *Handbook of Algorithms and Data Structures*. Addison-Wesley, Wokingham, England, 2nd edition, 1991.

[311] G. Gonnet, R. Baeza-Yates, and T. Snider. New indices for text: Pat trees and Pat arrays. In William Frakes and Ricardo Baeza-Yates, editors, *Information Retrieval: Data Structures and Algorithms*, pages 66–82. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[312] J. Gordon and E. Shortliffe. The Dempster-Shafer Theory of Evidence. In *Rule-Based Expert Systems*, pages 113–138. Addison-Wesley Publishing Company, 1984.

[313] Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 37th Allerton Institute 1995, January 1996. http://edfu.lis.uiuc.edu/allerton/95/.

[314] Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 38th Allerton Institute, January 1997. http://edfu.lis.uiuc.edu/allerton/96/.

[315] P. Graham. Glossary on Digital Library Terminology, 1997. Informal file sent by electronic mail for comments, available with permission of the author if suitable attribution is made.

[316] L. Gravano, K. Chang, H. García-Molina, C. Lagoze, and A. Paepcke. STARTS: Stanford protocol proposal for Internet retrieval and search. Technical report, Stanford University, Digital Library Project, 1997. http://www-db.stanford.edu/~gravano/starts.html.

[317] Luis Gravano, Chen-Chuan K. Chang, and Hector García-Molina. STARTS: Stanford proposal for Internet meta-searching. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, pages 207–218, Tucson, AZ, May 1997.

[318] Luis Gravano, Hector García-Molina, and Anthony Tomasic. The effectiveness of GlOSS for the text-database discovery problem. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, pages 126–137, Minneapolis, MN, May 1994.

[319] Luis Gravano, Hector García-Molina, and Anthony Tomasic. GlOSS: Text-source discovery over the Internet. *ACM Transactions on Database Systems*, 1999. To appear.

[320] S. J. Green. Automated link generation: can we do better than term repetition. In *7th WWW Conference*, Brisbane, Australia, 1998.

[321] Stephan Green, Gary Marchionini, Catherine Plaisant, and Ben Shneiderman. Previews and overviews in digital libraries: Designing surrogates to support visual information seeking. Technical Report Department of Computer Science CS-TR-3838, University of Maryland, 1997.

[322] Sharon L. Greene, Susan J. Devlin, Philip E. Cannata, and Louis M. Gomez. No IFs, ANDs, or ORs: A study of database querying. *International Journal of Man-Machine Studies*, 32(3):303–326, 1990.

[323] Gregory Grefenstette. *Cross-Language Information Retrieval.* Kluwer Academic Publishers, Boston, USA, 1998.

[324] S. M. Griffin. NSF/DARPA/NASA Digital Libraries Initiative: A program manager's perspective. *D-Lib Magazine*, July/August 1998. http://www.dlib.org/dlib/july98/07griffin.html.

[325] D. A. Grossman, O. Frieder, D. O. Holmes, and D. C. Roberts. Integrating structured data and text: A relational approach. *Journal of the American Society for Information Science*, 48:122–132, 1997.

[326] David A. Grossman and Ophir Frieder. *Information Retrieval: Algorithms and Heuristics.* Kluwer Academic Publishers, 1998.

[327] V. Gudivada, V. Raghavan, W. Grosky, and R. Kasanagottu. Information retrieval on the World Wide Web. *IEEE Internet Computing*, Oct-Nov:58–68, 1997.

[328] J.-C. Guédon. The virtual library: An oxymoron? NLM and MLA 1998 Leiter Lecture, National Library of Medicine, Bethesda, MD, May 1998.

[329] Catherine Guinan and Alan F. Smeaton. Information retrieval from hypertext using dynamically planned guided tours. In *Proc. of the 4th ACM Conference on Hypertext*, pages 122–130, 1992.

[330] A. Guttman. R-trees: A dynamic index structure for spatial searching. In *Proc. ACM SIGMOD*, pages 47–57, Boston, Mass, June 1984.

[331] Trudi Bellardo Hahn. Pioneers of the online age. *Information Processing & Management*, 32:33–48, 1996.

[332] David Haines and W. Bruce Croft. Relevance feedback and inference networks. In *Proc. of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2–11, Pittsburgh, PA, USA, 1993.

[333] J. Hakala. The 5th Dublin Core Metadata Workshop, October 1997. http://linnea.helsinki.fi/meta/DC5.html.

[334] Frank G. Halasz, Thomas P. Moran, and Randall H. Trigg. Notecards in a nutshell. In *Proc. of ACM CHI+GI Conference on Human Factors in Computing Systems and Graphics Interface*, pages 45–52, 1987.

[335] Micheline Hancock-Beaulieu. User friendliness and human-computer interaction in online library catalogs. *Program*, 46(1):29–37, 1992.

[336] Micheline Hancock-Beaulieu. Experiments on interfaces to support query expansion. *Journal of Documentation*, 53(1):8–19, 1997.

[337] Micheline Hancock-Beaulieu, Margaret Fieldhouse, and Thien Do. An evaluation of interactive query expansion in an online library catalogue with a graphical user interface. *Journal of Documentation*, 51(3):225–243, 1995.

[338] Micheline Hancock-Beaulieu and Stephen Walker. An evaluation of automatic query expansion in an online library catalogue. *Journal of Documentation*, 48(4):406–421, 1992.

[339] D. Harman. Ranking algorithms. In W. B. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures & Algorithms*, pages 363–392. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[340] D. Harman. Relevance feedback and other query modification techniques. In W. B. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures & Algorithms*, pages 241–263. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[341] D. Harman, E. Fox, R. Baeza-Yates, and W. Lee. Inverted files. In W. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Algorithms and Data Structures*, chapter 3, pages 28–43. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[342] D. K. Harman. Overview of the third text retrieval conference. In *Proc. of the 3rd Text REtrieval Conference (TREC-3)*, pages 1–19, Gaithersburg, USA, 1995. National Institute of Standards and Technology Special Publication.

[343] Donna Harman. Relevance feedback revisited. In *Proc. of the 5th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1–10, Copenhagen, Denmark, 1992.

[344] D. J. Harper. *Relevance Feedback in Document Retrieval Systems: An Evaluation of Probabilistic Strategies*. PhD thesis, Jesus College, Cambridge, England, 1980.

[345] D. J. Harper and C. J. van Rijsbergen. An evaluation of feedback in document retrieval using co-occurrence data. *Journal of Documentation*, 34(3):189–216, 1978.

[346] G. Harper. The Conference on Fair Use (CONFU), September 1997. http://www.utsystem.edu/ogc/intellectualproperty/confu.htm.

[347] S. P. Harter. What is a digital library? Definitions, content, and issues. In *Proc. of KOLISS DL '96: International Conference on Digital Libraries and Information Services for the 21st Century*, Seoul, Korea, September 1996. http://php.indiana.edu/~harter/korea-paper.htm.

[348] Stephen P. Harter. *Online Information Retrieval.* Academic Press, 1986.

[349] S. Harum. Digital Library Initiative, January 1998. http://dli.grainger.uiuc.edu/national.htm.

[350] Alexander G. Hauptmann and Bert F. Green. A comparison of command, menu-selection and natural-language computer programs. *Behaviour and Information Technology*, 2(2):163–178, 1983.

[351] Leigh Watson Healy. *Library Systems: Current Developments and Future Directions.* Council on Library and Information Resources, Washington DC, 1998.

[352] J. Heaps. *Information Retrieval — Computational and Theoretical Aspects.* Academic Press, 1978.

[353] M. Hearst, G. Kopec, and D. Brotsky. Research in support of digital libraries at Xerox PARC, Part II: Paper and digital documents. *D-Lib Magazine*, June 1996. http://www.dlib.org/dlib/june96/hearst/06hearst.html.

[354] M. A. Hearst. Research in Support of Digital Libraries at Xerox PARC, Part I: The changing social roles of documents. *D-Lib Magazine*, May 1996. http://www.dlib.org/dlib/may96/05hearst.html.

[355] Marti A. Hearst. TileBars: Visualization of term distribution information in full text information access. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 59–66, Denver, CO, May 1995.

[356] Marti A. Hearst. Improving full-text precision using simple query constraints. In *Proc. of the 5th Annual Symposium on Document Analysis and Information Retrieval*, Las Vegas, NV, 1996.

[357] Marti A. Hearst. The use of categories and clusters in organizing retrieval results. In Tomek Strzalkowski, editor, *Natural Language Information Retrieval.* Kluwer Academic Publishers, 1999. To appear.

[358] Marti A. Hearst and Chandu Karadi. Cat-a-cone: An interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. In *Proc. of the 20th Annual International ACM SIGIR Conference*, pages 246–255, Philadelphia, PA, 1997.

[359] Marti A. Hearst and Jan O. Pedersen. Reexamining the cluster hypothesis: Scatter/gather on retrieval results. In *Proc. of the 19th Annual International ACM SIGIR Conference*, pages 76–84, Zurich, Switzerland, 1996.

[360] L. Heath, D. Hix, L. Nowell, W. Wake, G. Averboch, and E. Fox. Envision: A user-centered database from the computer science literature. *Communications of the ACM*, 38(4):52–53, April 1995.

[361] J. H. Hellerstein and M. Stonebraker. Predicate migration: Optimizing queries with expensive predicates. In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, pages 267–276, Washington, DC, USA, 1993.

[362] Joseph M. Hellerstein, Jeffrey F. Naughton, and Avi Pfeffer. Generalized search trees for database systems. In *Proc. of VLDB Conf.*, pages 562–573, Zurich, Switzerland, Sept 1995.

[363] Matthias Hemmje, Clemens Kunkel, and Alexander Willett. LyberWorld – a visualization user interface supporting fulltext retrieval. In *Proc. of the 17th Annual International ACM SIGIR Conference*, pages 249–259, Dublin, Ireland, July 1994.

[364] R. Hendly, N. Drew, A. Wood, and R. Beale. Narcissus: Visualizing information. In *Proc. of the IEEE Information Visualization Symposium*, pages 90–96, Atlanta, GA, USA, Oct 1995.

[365] David G. Hendry and David J. Harper. An informal information-seeking environment. *Journal of the American Society for Information Science*, 48(11):1036–1048, 1997.

[366] M. Hersovici, M. Jacobi, Y. S. Maarek, D. Pelleg, M. Shtalhaim, and S. Ur. The shark-search algorithm. An application: tailored Web site mapping. In *7th WWW Conference*, Brisbane, Australia, April 1998.

[367] C. Hert. Information retrieval: A social informatics perspective. Allerton discussion document submitted for the 1996 Allerton Institute, 1996.

[368] Morten Hertzum and Erik Frokjaer. Browsing and querying in online documentation: A study of user interfaces and the interaction process. *ACM Transactions on Computer-Human Interaction*, 3(2):136–161, 1996.

[369] Eric van Herwijnen. *Practical SGML*. Kluwer Academic Publishers, 2nd edition, 1994.

[370] Ron R. Hightower, Laura T. Ring, Jonathan I. Helfman, Benjamin B. Bederson, and James D. Hollan. Graphical multiscale Web histories: A study of padprints. In *Proc. of the Ninth ACM Conference on Hypertext*, pages 58–65, Pittsburgh, PA, USA, 1998.

[371] Charles R. Hildreth. The use and understanding of keyword searching in an university online catalog. *Information Technology and Libraries*, 16:52–62, 1997.

[372] C. R. Hildreth. Online library catalogues as IR systems: what can we learn from research? In P. A. Yates Mercer, editor, *Future Trends in Information Science and Technology*, pages 9–25. Taylor-Graham, Los Angeles, 1988.

[373] Rainer Himmeroder, Georg Lausen, Bertram Ludascher, and Christian Schlepphorst. On a declarative semantics for Web queries. In *Proc. of the Int. Conf. on Deductive and Object-Oriented Database (DOOD)*, pages 386–398, Singapore, December 1997.

[374] D. S. Hirschberg and D. A. Lelewer. Efficient decoding of prefix codes. *Communications of the ACM*, 33(4), 1990.

[375] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley Publishing Company, 1979.

[376] R. N. Horspool. Practical fast searching in strings. *Software Practice and Experience*, 10:501–506, 1980.

[377] R. N. Horspool and G. V. Cormack. Constructing word-based text compression algorithms. In *Proc. of IEEE Second Data Compression Conference*, pages 62–81, 1992.

[378] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proc. of AAAI 98*, Madison, WI, July 1998.

[379] P. Hoschka. Synchronized multimedia integration language. W3C Working Draft, February 1998. http://www.w3.org/TR/WD-smil.

[380] Hotbot: Main page. http://www.hotbot.com, 1996.

[381] T. Y. Hou, A. Hsu, P. Liu, and M. Y. Chiu. A content-based indexing technique using relative geometry features. In *Proc. of SPIE'92*, volume 1662, pages 59–68, 1992.

[382] N. Van House, D. Levy, A. Bishop, and B. Buttenfield. User needs assessment and evaluation: issues and methods (workshop). In *Proc. of the 1st ACM International Conference on Digital Libraries*, page 186, College Station, TX, USA, 1996.

[383] Adele Howe and Danielle Dreilinger. Savvysearch: A metasearch engine that learns which search engines to query. *AI Magazine*, 18(2):19–25, 1997.

[384] J-Y. Huang, C-T. Fang-Tsou, and J-L. Chang. Multiuser 3D Web browsing system. *IEEE Internet Computing*, 2(5):70–73, 1998.

[385] M. L. Huang, P. Eades, and R. F. Cohen. WebOFDAV — navigating and visualizing the Web on-line with animated context swapping. In *7th WWW Conference*, Brisbane, Australia, April 1998.

[386] D. Huffman. A method for the construction of minimum-redundancy codes. *Proc. of the I.R.E.*, 40(9):1090--1101, 1952.

[387] D. A. Hull and G. Grefenstette. Querying across languages: A dictionary-based approach to multilingual information retrieval. In *Proc. of the 19th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 49–57, Zurich, Switzerland, 1996.

[388] David A. Hull, Jan O. Pedersen, and Hinrich Schütze. Method combination for document filtering. In *Proc. of the 19th Annual Int. ACM SIGIR Conference*, pages 279–287, Zurich, Switzerland, 1996.

[389] R. Iannella. Australian digital library initiatives. *D-Lib Magazine*, December 1996. http://www.dlib.org/dlib/december96/12iannella.html.

[390] IBM Corporation, IBM Digital Library, 1998. http://www.software.ibm.com/is/diglib/.

[391] E. Ide. New experiments in relevance feedback. In G. Salton, editor, *The SMART Retrieval System*, pages 337–354. Prentice Hall, 1971.

[392] IEEE Standards Committee on Optical Disk and Multimedia Platforms (SC ODMP), IEEE SFQL. Technical report, IEEE, Washington, USA, 1992.

[393] G. H. Brett II. An integrated system for distributed information services. *D-Lib Magazine*, December 1996. http://www.dlib.org/dlib/december96/dipps/12brett.html.

[394] Infoseek Corporation, Distributed Search Patent. http://software.infoseek.com/-patents/dist_search/Default.htm.

[395] Inktomi: Main Page. http://www.inktomi.com, 1998.

[396] International DOI Foundation, Digital Object Identifier System, June 1998. http://www.doi.org/index.html.

[397] H. Ishikawa, F. Suzuki, F. Kozakura, A. Makinouchi, M. Miyagishima, M. Aoshima Y. Izumida, and Y. Yamane. The model, language and implementation of an object-oriented multimedia knowledge base management system. *ACM Transactions on Database Systems*, 18:1–50, 1993.

[398] ISO. *Office Document Architecture (ODA) an Interchange Format*, 1986.

[399] Paul S. Jacobs and Lisa F. Rau. Innovations in text interpretation. *Artificial Intelligence*, 63(1-2):143–191, 1993.

[400] H. V. Jagadish. A retrieval technique for similar shapes. In *Proc. ACM SIGMOD Conf.*, pages 208–217, Denver, Colorado, May 1991.

[401] Jango: Main Page. http://www.jango.com, 1998.

[402] M. Jarke and J. Koch. Query optimization in database systems. *ACM Computing Surveys*, 16(2):111–152, June 1984.

[403] B. J. Jensen, A. Spink, J. Bateman, and T. Saracevic. Real life information retrieval: A study of user queries on the Web. *ACM SIGIR Forum*, 32(1):5–17, 1998.

[404] Byeong-Soo Jeong and Edward Omiecinski. Inverted file partitioning schemes in multiple disk systems. *IEEE Trans. Par. and Dist. Syst.*, 6(2):142–153, February 1995.

[405] H. Jiang, A. K. Elmagarmid, A. A. Helal, A. Joshi, and M. Ahmed. *Video Database Systems*. Kluwer Academic, San Francisco, CA, 1997.

[406] Thorsten Joachims, Dayne Freitag, and Tom Mitchell. WebWatcher: A tour guide for the World Wide Web. In *Proc. of the 15th Int. Joint Conference on Artificial Intelligence*, Nagoya, Japan, August 1997.

[407] Brian Johnson and Ben Shneiderman. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In *Proc. of the 2nd Int. IEEE Visualization Conference*, pages 284–291, San Diego, USA, 1991.

[408] I. T. Jolliffe. *Principal Component Analysis*. Springer Verlag, 1986.

[409] K. Sparck Jones. A statistical interpretation of term specificity and its application to retrieval. *Journal of Documentation*, 28(1):11–20, 1972.

[410] K. Sparck Jones. A statistical interpretation of term specificity and its application to retrieval. *Information Storage and Retrieval*, 9(11):619–633, 1973.

[411] K. Sparck Jones. Experiments in relevance weighting of search terms. *Information Processing & Management*, 15(13):133–144, 1979.

[412] K. Sparck Jones. Search term relevance weighting given little relevance information. *Journal of Documentation*, 35(1):30–48, 1979.

[413] K. Sparck Jones and E. O. Barber. What makes an automatic keyword classification effective. *Journal of the American Society for Information Sciences*, 22(3):166–175, 1971.

[414] K. Sparck Jones and P. Willet. *Readings in Information Retrieval*. Morgan Kaufmann Publishers, Inc., 1997.

[415] Karen Sparck Jones. The Cranfield tests. In Karen Sparck Jones, editor, *Information Retrieval Experiment*, pages 256–284. Butterworths, London, 1981.

[416] S. Jones, S. Walker, M. Gatford, and T. Do. Peeling the onion: Okapi system architecture and software design issues. *Journal of Documentation*, 53:58–68, 1997.

[417] Steve Jones. Graphical query specification and dynamic result previews for a digital library. In *Proc. of UIST'98, ACM Symposium on User Interface Software and Technology*, San Francisco, USA, November 1998.

[418] T. Joseph and A. F. Cárdenas. PICQUERY: A high level language for pictorial database management. *IEEE Transactions on Software Engineering*, 14(5):630–638, 1988.

[419] T. Joyce and R. M. Needham. The thesaurus approach to information retrieval. In K. Sparck Jones and P. Willet, editors, *Readings in Information Retrieval*, pages 15–20. Morgan Kaufmann Publishers, Inc., 1997.

[420] D. Austin Henderson Jr and Stuart K. Card. Rooms: The use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics*, 5(3):211–243, 1986.

[421] E. J. Whitehead Jr and M. Wiggins. WebDAV: IETF standard for collaborative authoring on the Web. *IEEE Internet Computing*, 2(5):34–40, 1998.

[422] W. M. Shaw Jr, R. Burgin, and P. Howell. Performance standards and evaluations in IR test collections: Cluster-based retrieval models. *Information Processing & Management*, 33(1):1–14, 1997.

[423] W. M. Shaw Jr., R. Burgin, and P. Howell. Performance standards and evaluations in IR test collections: Vector-space and other retrieval models. *Information Processing & Management*, 33(1):15–36, 1997.

[424] B. Kahle. Archiving the Internet. http://www.alexa.com/~brewster/essays/sciam_article.html, 1997.

[425] B. Kahle and A. Medlar. An information server for corporate users: Wide Area Information Servers. *ConneXions — The Interoperability Report*, 5(11):2–9, 1991. ftp://think.com/wais/wais-corporate-paper.text.

[426] R. Kahn and R. Wilensky. A framework for distributed digital object services. Technical Report cnri.dlib/tn95-01, CNRI, May 1995. http://www.cnri.reston.va.us/k-w.html.

[427] Ibrahim Kamel and Christos Faloutsos. Hilbert R-tree: An improved R-tree using fractals. In *Proc. of VLDB Conference*, pages 500–509, Santiago, Chile, Sept 1994.

[428] Eser Kandogan and Ben Shneiderman. Elastic windows: Evaluation of multi-window operations. In *Proc. of ACM Conference on Human Factors in Computing Systems*, volume 1, pages 250–257, Atlanta, GA, USA, March 1997.

[429] P. B. Kantor. Assessing the factors leading to adoption of digital libraries, and growth in their impacts: the Goldilocks principle. Allerton discussion document submitted for the 1996 Allerton Institute. http://edfu.lis.uiuc.edu/allerton/96/kantor.html.

[430] J. Karkkäinen and E. Ukkonen. Two and higher dimensional pattern matching in optimal expected time. In *Proc. of ACM-SIAM Symposium on Discrete Algorithms*, pages 715–723, San Francisco, USA, 1994.

[431] Norio Katayama and Shin'ichi Satoh. The SR-tree: An index structure for high-dimensional nearest neighbor queries. In *Proc. of ACM SIGMOD*, pages 369–380, Tucson, AZ, 1997.

[432] Henry Kautz, Bart Selman, and Mehul Shah. The hidden Web. *AI Magazine*, 18(2):27–36, 1997.

[433] P. R. Keller and M. M. Keller. *Visual Cues: Practical Data Visualization*. IEEE Computer Society Press, 1993.

[434] R. Kengeri, C. D. Seals, H. P. Reddy, H. D. Harley, and E. A. Fox. Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD, December 1997. http://fox.cs.vt.edu/~fox/u/Usability.pdf.

[435] M. M. Kessler. Comparison of results of bibliographic coupling and analytic subject indexing. *American Documentation*, 16(3):223–233, July 1965.

[436] R. Khare and A. Rifkin. XML: A door to automated Web applications. *IEEE Internet Computing*, 1(4):78–86, 1977.

[437] R. Khare and A. Rifkin. The origin of (document) species. In *7th WWW Conference*, Brisbane, Australia, April 1998.

[438] S. Khoshafian and A. B. Baker. *Multimedia and Imaging Databases*. Morgan Kauffman, 1996.

[439] P. Kilpelainen and H. Mannila. Retrieval from hierarchical texts by partial patterns. In *Proc. of the 16th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 214–222, Pittsburgh, USA, 1993.

[440] Hanhwe Kim and Stephen C. Hirtle. Spatial metaphors and disorientation in hypertext browsing. *Behaviour and Information Technology*, 14(4):239–250, 1995.

[441] Steven T. Kirsch. Document retrieval over networks wherein ranking and relevance scores are computed at the client for multiple database documents. US Patent 5,659,732, August 1997.

[442] W. Klas and A. Sheth. Special issue on meta-data for digital media. *SIGMOD Record*, 23(4):19–20, 1994.

[443] Adrienee J. Kleiboemer, Manette B. Lazear, and Jan O. Pedersen. Tailoring a retrieval system for naive users. In *Proc. of the 5th Annual Symposium on Document Analysis and Information Retrieval*, Las Vegas, NV, USA, 1996.

[444] Jon Kleinberg. Authoritative sources in a hyperlinked environment. In *Proc. of the 9th ACM-SIAM Symposium on Discrete Algorithms*, pages 668–677, San Francisco, USA, Jan 1998.

[445] D. E. Knuth. *The Art of Computer Programming*, volume 3: Searching and Sorting. Addison-Wesley, 1973.

[446] D. E. Knuth. Dynamic Huffman coding. *Journal of Algorithms*, 6:163–180, 1985.

[447] D. E. Knuth, J. H. Morris, Jr, and V. R. Pratt. Fast pattern matching in strings. *SIAM Journal on Computing*, 6(1):323–350, 1977.

[448] Jurgen Koenemann and Nicholas J. Belkin. A case for interaction: A study of interactive information retrieval behavior and effectiveness. In *Proc. of ACM Conference on Human Factors in Computing Systems*, volume 1 of *Papers*, pages 205–212, Zurich, Switzerland, 1996.

[449] Janet L. Kolodner. *Case-based Reasoning.* Morgan Kaufmann Publishers, 1993.

[450] D. Konopnicki and O. Shmueli. W3QS: A query system for the World Wide Web. In *Proc. of VLDB'95*, pages 54–65, Zurich, Switzerland, September 1995.

[451] Robert Korfhage. *Information Storage and Retrieval.* John Wiley & Sons, Inc., 1997.

[452] Robert R. Korfhage. To see or not to see – is that the query? In *Proc. of the 14th Annual Int. ACM SIGIR Conference*, pages 134–141, Chicago, USA, 1991.

[453] Flip Korn and Ben Shneiderman. Navigating terminology hierarchies to access a digital library of medical images. Technical Report HCIL-TR-94-03, University of Maryland, USA, Korn & Shneiderman, 1995.

[454] Flip Korn, Nikolaos Sidiropoulos, Christos Faloutsos, Eliot Siegel, and Zenon Protopapas. Fast nearest-neighbor search in medical image databases. *Proc. Conf. on Very Large Data Bases*, pages 215–226, September 1996.

[455] J. Korpela. Lurching toward Babel: HTML, CSS, and XML. *IEEE Computer*, 31(7):103–106, 1998.

[456] S. M. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3:185–226, 1989.

[457] M. Koster. Guidelines for robot writers, 1993. http://info.webcrawler.com/mak/-projects/robots/guidelines.html.

[458] Nick Koudas, Christos Faloutsos, and Ibrahim Kamel. Declustering spatial databases on a multi-computer architecture. *EDBT Conf. Proc.*, pages 592–614, March 1996.

[459] Gerald Kowalski. *Information Retrieval Systems, Theory and Implementation.* Kluwer Academic Publishers, Boston, USA, 1997.

[460] Robyn Kozierok and Pattie Maes. A learning interface agent for scheduling meetings. In *Proc. of the 1993 Int. Workshop on Intelligent User Interfaces*, pages 81–88, New York, NY, 1993.

[461] D. Kraft and D. A. Buel. Fuzzy sets and generalized Boolean retrieval systems. *International Journal of Man-Machine Studies*, 19:45–56, 1983.

[462] Joseph B. Kruskal and Myron Wish. *Multidimensional scaling.* SAGE publications, Beverly Hills, 1978.

[463] Julian Kupiec. MURAX: A robust linguistic approach for question answering using an on-line encyclopedia. In *Proc. of the 16th Annual Int. ACM SIGIR Conference*, pages 181–190, Pittsburgh, PA, 1993.

[464] Julian Kupiec, Jan Pedersen, and Francine Chen. A trainable document summarizer. In *Proc. of the 18th Annual Int. ACM SIGIR Conference*, pages 68–73, Seattle, WA, 1995.

[465] K. L. Kwok, L. Grunfeld, and D. D. Lewis. TREC-3 ad-hoc, routing retrieval and thresholding experiments using PIRCS. In *Proc. of the Text REtrieval Conference*, pages 247–256, Gaithersburg, MD, USA, 1995.

[466] K. L. Kwok. A neural network for probabilistic information retrieval. In *Proc. of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 21–30, Cambridge, USA, June 1989.

[467] K. L. Kwok. Experiments with a component theory of probabilistic information retrieval based on single terms as document components. *ACM Transactions on Information Systems*, 8(4):363–386, October 1990.

[468] K. L. Kwok. A network approach to probabilistic information retrieval. *ACM Transactions on Information Systems*, 13(3):324–353, July 1995.

[469] K. L. Kwok, L. Papadopolous, and Y. Y. Kwan. Retrieval experiments with a large collection using PIRCS. In *Proc. of the 1st TExt Retrieval Conference*, pages 153–172, Gaithersburg, MD, USA, 1993. Special Publication 500-267, National Institute of Standards and Technology (NIST).

[470] Eric Lagergren and Paul Over. Comparing interactive information retrieval systems across sites: The TREC-6 interactive track matrix experiment. In *Proc. of the 21st Annual Int. ACM SIGIR Conference*, pages 164–172, Melbourne, Australia, 1998.

[471] C. Lagoze. Networked Computer Science Technical Reference Library. http://www.ncstrl.org.

[472] C. Lagoze. The Warwick framework: a container architecture for diverse sets of metadata. *D-Lib Magazine*, July/August 1996. http://www.dlib.org/dlib/july96/lagoze/07lagoze.html.

[473] C. Lagoze and D. Ely. Implementation issues in an open architecture framework for digital object services. Technical Report TR95-1540, Cornell University Computer Science Department, 1995.

[474] C. Lagoze, D. Fielding, and S. Payette. Making global digital libraries work: collection services, connectivity regions, and collection views. In *Proc. of the 3rd ACM International Conference on Digital Libraries*, 1998.

[475] C. Lagoze, R. McGrath, E. Overly, and N. Yeager. A design for inter-operable secure object stores (ISOS). Technical Report TR95-1558, Cornell University Computer Science Department, 1995.

[476] Laks V. S. Lakshmanan, Fereidoon Sadri, and Iyer N. Subramanian. A declarative language for querying and restructuring the Web. In *6th Int. Workshop on Research Issues in Data Engineering, RIDE '96*, New Orleans, February 1996.

[477] B. LaMacchia. The Internet fish construction kit. In *6th. Int'l. WWW Conference*, Santa Clara, CA, USA, April 1997.

[478] John Lamping, Ramana Rao, and Peter Pirolli. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proc. of the ACM Conference on Human Factors in Computing Systems*, pages 401–408, Denver, CO, USA, May 1995.

[479] F. W. Lancaster. Natural language in information retrieval. In *Indexing and Abstracting in Theory and Practice*, pages 193–218. University of Illinois, Champaign, IL, 1991.

[480] G. Landau and U. Vishkin. Fast string matching with k differences. *Journal of Computer Systems Science*, 37:63–78, 1988.

[481] Thomas K. Landauer, Dennis E. Egan, Joel R. Remde, Michael Lesk, Carol C. Lochbaum, and Daniel Ketchum. Enhancing the usability of text through computer delivery and formative evaluation: the SuperBook project. In C. McKnight, A. Dillon, and J. Richardson, editors, *Hypertext: A Psychological Perspective*, pages 71–136. Ellis Horwood, 1993.

[482] Andrew Large and Jamshid Beheshti. OPACS: a research review. *LISR*, 19:111–133, 1997.

[483] Jill H. Larkin and Herbert A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11:65–99, 1987.

[484] Ray R. Larson. Between Scylla and Charybdis: Subject searching in the online catalog. *Advances in Librarianship*, 15:175–236, 1991.

[485] Ray R. Larson. Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace. In *Proc. of the 1996 Annual ASIS Meeting*, pages 71–78, 1996.

[486] Ray R. Larson, Jerome McDonough, Paul O'Leary, and Lucy Kuntz. Cheshire II: Designing a next-generation online catalog. *Journal of the American Society for Information Science*, 47(7):555–567, 1996.

[487] Ora Lassila. Web metadata: A matter of semantics. *IEEE Internet Computing*, 2(4):30–37, 1998.

[488] S. Lawrence and C. L. Giles. Inquirus, the NECI meta search engine. In *7th WWW Conference*, pages 95–105, Brisbane, Australia, 1998.

[489] S. Lawrence and C. Lee Giles. Context and page analysis for improved Web search. *IEEE Internet Computing*, 2(4):38–46, 1998.

[490] S. Lawrence and C. L. Giles. Searching the World Wide Web (in reports). *Science*, 280(5360):98, April 3 1998.

[491] Doug Lea. *Concurrent Programming in Java: Design Principles and Patterns*. The Java Series. Addison-Wesley, Reading, MA, 1997.

[492] David B. Leake, editor. *Case-based Reasoning: Experiences, Lessons, and Future Directions*. AAAI Press, Menlo Park, CA, 1996.

[493] J. H. Lee, W. Y. Kim, and Y. H. Lee. Ranking documents in thesaurus-based Boolean retrieval systems. *Information Processing & Management*, 30(1):79–91, 1993.

[494] J. J. Lee and P. Kantor. A study of probabilistic information retrieval systems in the case of inconsistent expert judgements. *Journal of the American Society for Information Sciences*, 42(3):166–172, 1991.

[495] Joon Ho Lee. Properties of extended Boolean models in information retrieval. In *Proc. of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 182–190, Dublin, Ireland, 1994.

[496] Joon Ho Lee, Won Yong Kim, Myoung Ho Kim, and Yoon Joon Lee. On the evaluation of Boolean operators in the extended Boolean retrieval framework. In *Proc. of the 16th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 291–297, Pittsburgh, PA, USA, 1993.

[497] O. Lehtinen, E. Sutinen, and J. Tarhio. Experiments on block indexing. In *Proc. of WSP'96*, pages 183–193, Recife, Brazil, 1996. Carleton University Press.

[498] H. V. Leighton and J. Srivastava. Precision among World Wide Web search engines: AltaVista, Excite, Hotbot, Infoseek, and Lycos. http://www.winona.msus.edu/library/webind2/webind2.htm, 1997.

[499] B. Leiner. D-Lib Working Group on Digital Library Metrics, May 1998. http://www.dlib.org/metrics/public/metrics-home.html.

[500] M. E. Lesk. Word-word associations in document retrieval systems. *American Documentation*, 20(1):8–36, 1969.

[501] Michael Lesk. *Practical Digital Libraries; Books, Bytes, and Bucks*. Morgan Kaufmann, 1997.

[502] Y. K. Leung and M. D. Apperley. A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction*, 1(2):126–160, 1994.

[503] R. LeVan. Dublin Core and Z39.50. Draft version 1.2, February 1998. http://cypress.dev.oclc.org:12345/~rrl/docs/dublincoreandz3950.html.

[504] V. Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966. Original in Russian in *Doklady Akademii Nauk SSSR, 163(4):8 45–848, 1965.*

[505] D. M. Levy. Topics in document research. In *Proc. of the ACM Conference on Document Processing Systems (SIGDOC '88)*, pages 187–193, 1988.

[506] D. M. Levy. Fixed or fluid?: document stability and new media. In *Proc. of the 1994 ACM European Conference on Hypermedia Technology*, pages 24–31, 1994.

[507] D. M. Levy. I read the news today, oh boy: reading and attention in digital libraries. In *Proc. of the 2nd ACM Int. Conference on Digital Libraries*, pages 202–211, Philadelphia, PA, USA, July 1997.

[508] D. M. Levy, D. C. Brotsky, and K. R. Olson. Formalizing the figural: aspects of a foundation for document manipulation. In *Proc. of the ACM Conference on Document Processing Systems (SIGDOC '88)*, pages 145–151, 1988.

[509] D. M. Levy and C. C. Marshall. Going digital: a look at assumptions underlying digital libraries. *Communications of the ACM*, 38:77–84, April 1995.

[510] The LEXIS-NEXIS Computing Complex. http://www.lexis-nexis.com/lncc/about/datacenter.html.

[511] W-S. Li, J. Shim, K. S. Candan, and Y. Hara. WebDB: A Web query system and its modeling, language, and implementation. In *Proc. of Advances in Digital Libraries*, Santa Barbara, CA, USA, April 1998.

[512] Y. Li. Toward a qualitative search engine. *IEEE Internet Computing*, 2(4):24–29, 1998.

[513] Library of Congress, Metadata, Dublin Core and USMARC: A Review of Current Efforts, January 1997. gopher://marvel.loc.gov/00/.listarch/usmarc/dp99.doc.

[514] Library of Congress, MARC Standards, June 1998. http://lcweb.loc.gov/marc/marc.html.

[515] Library of Congress, Z39.50 Maintenance Agency, June 1998. http://lcweb.loc.gov/z3950/agency/.

[516] J. C. R. Licklider. *Libraries of the Future*. Cambridge, MA.: M.I.T. Press, 1965.

[517] H. W. Lie and B. Bos. *Cascading Style Sheets: Designing for the Web*. Addison-Wesley, 1997.

[518] Henry Lieberman. Letizia: an agent that assists Web browsing. In *Proc. of the 14th Int. Joint Conference on Artificial Intelligence*, pages 924–929, 1995.

[519] King-Ip Lin, H. V. Jagadish, and Christos Faloutsos. The TV-tree — an index structure for high-dimensional data. *VLDB Journal*, 3:517–542, October 1994.

[520] Xia Lin, Dagobert Soergel, and Gary Marchionini. A self-organizing semantic map for information retrieval. In *Proc. of the 14th Annual Int. ACM SIGIR Conference*, pages 262–269, Chicago, 1991.

[521] Ming-Ling Lo and Chinya V. Ravishankar. Spatial joins using seeded trees. In *Proc. of ACM SIGMOD*, pages 209–220, Minneapolis, MN, USA, May 1994.

[522] R. M. Losee and A. Bookstein. Integrating Boolean queries in conjunctive normal form with probabilistic retrieval models. *Information Processing & Management*, 24(3):315–321, 1988.

[523] Henry J. Lowe and G. Octo Barnett. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *Journal of the American Medical Association*, 271(4):1103–1108, 1994.

[524] Zhihong Lu, Kathryn S. McKinley, and Brendon Cahoon. The hardware/software balancing act for information retrieval on symmetric multiprocessors. Technical Report TR98-25, Dept. of Comp. Sci., Univ. of Mass., Amherst, MA, 1998.

[525] C. Lynch and H. García-Molina. Interoperability, scaling, and the digital libraries research agenda: A report on the May 18-19, 1995, IITA Digital Libraries Workshop, August 1995. http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html.

[526] Clifford Lynch. The next generation of public access information retrieval systems for research libraries: lessons from ten years of the MELVYL system. *Information Technology and Libraries*, 11(4):405–415, 1992.

[527] Y. Maarek, M. Jacovi, M. Shtalhaim, S. Ur, D. Zernik, and I. Z. Ben Shaul. WebCutter: A system for dynamic and tailorable site mapping. In *6th WWW Conference*, pages 713–722, Santa Clara, CA, USA, 1997.

[528] Y. S. Maarek. Organizing documents to support browsing in digital libraries. Allerton discussion document submitted for the 1995 Allerton Institute. http://edfu.lis.uiuc.edu/allerton/95/s4/maarek.html.

[529] Y. S. Maarek and A. J. Wecker. The librarian's assistant: Automatically assembling books into dynamic bookshelves. In *Proc. of RIAO '94: Intelligent Multimedia Information Retrieval Systems and Management*, New York, USA, October 1994.

[530] Jock Mackinlay, George Robertson, and Stuart K. Card. The perspective wall: Detail and context smoothly integrated. In *Proc. of the ACM Conference on Human Factors in Computing Systems*, pages 173–179, 1991.

[531] Jock D. Mackinlay, Ramana Rao, and Stuart K. Card. An organic user interface for searching citation links. In *Proc. of ACM Conference on Human Factors in Computing Systems*, volume 1 of *Papers*, pages 67–73, Denver, CO, USA, 1995.

[532] A. MacLean, P. J. Barnard, and M. D. Wilson. Evaluating the human interface of a data entry system: User choice and performance measures yield different tradeoff functions. In *Proc. of the HCI'85 Conference on People and Computers: Designing the Interface*, pages 172–185, 1985.

[533] I. A. MacLeod. Storage and retrieval of structured documents. *Information Processing & Management*, 26(2):197–208, 1990.

[534] I. A. MacLeod. A query language for retrieving information from hierarchic text structures. *The Computer Journal*, 34(3):254-264, 1991.

[535] Ian A. Macleod, T. Patrick Martin, Brent Nordin, and John R. Phillips. Strategies for building distributed information retrieval systems. *Information Processing & Management*, 23(6):511–528, 1987.

[536] Pattie Maes and Robyn Kozierok. Learning interface agents. In *Proc. of AAAI 93*, pages 459–465, Washington, DC, July 1993.

[537] U. Manber and P. Bigot. Search Broker: Main page. http://debussy.cs.arizona.edu/sb/, 1997.

[538] U. Manber and G. Myers. Suffix arrays: a new method for on-line string searches. In *Proc. of ACM-SIAM Symposium on Discrete Algorithms*, pages 319–327, San Francisco, USA, 1990.

[539] U. Manber, M. Smith, and B. Gopal. WebGlimpse: combining browsing and searching. In *Proc. of USENIX Technical Conference*, pages 195–206, Anaheim, USA, Jan 1997.

[540] Udi Manber and Sun Wu. GLIMPSE: A tool to search through entire file systems. In *Proc. of USENIX Technical Conference*, pages 23–32, San Francisco, USA, January 1994. ftp://cs.arizona.edu/glimpse/glimpse.ps.Z.

[541] B. Mandelbrot. *Fractal Geometry of Nature*. W. H. Freeman, New York, 1977.

[542] Gary Marchionini. *Information Seeking in Electronic Environments*. Cambridge University Press, Cambridge, 1995.

[543] M. Marchiori. The quest for correct information on the Web: Hyper search engines. In *6th WWW Conf.*, pages 265–274, Santa Clara, CA, USA, 1997.

[544] M. Marchiori. The limits of Web metadata and beyond. In *7th WWW Conference*, Brisbane, Australia, April 1998.

[545] S. Marcus and V. S. Subrahmanian. Foundations of Multimedia Database Systems. *Journal of the ACM*, 4(3):474–505, 1996.

[546] Karen Markey, Pauline Atherton, and Claudia Newton. An analysis of controlled vocabulary and free text search statements in online searches. *Online Review*, 4:225–236, 1982.

[547] M. E. Maron and J. L. Kuhns. On relevance, probabilistic indexing and information retrieval. *Association for Computing Machinery*, 7(3):216–244, 1960.

[548] C. C. Marshall. Annotation: from paper books to the digital library. In *Proc. of the 2nd ACM Int. Conference on Digital Libraries*, pages 131–141, Philadelphia, PA, USA, 1997.

[549] Y. Masunaga. Design issues of OMEGA an object-oriented multimedia database management system. *Journal of Information Processing*, 14:60–74, 1991.

[550] Mark T. Maybury. *Intelligent Multimedia Information Retrieval.* MIT Press, 1997.

[551] Ray McAleese, editor. *Hypertext I: Theory into Practice.* Ablex Publishing Corporation, 1988.

[552] B. McCune, R. Tong, J. S. Dean, and D. Shapiro. Rubric: A system for rule-based information retrieval. *IEEE Transactions on Software Engineering*, 11(9), 1985.

[553] Charles T. Meadow, Barbara A. Cerny, Christine L. Borgman, and Donald O. Case. Online access to knowledge: System design. *Journal of the American Society for Information Science*, 40(2):86–98, 1989.

[554] Daniel A. Menasce and Virgilio A. F. Almeida. *Capacity Planning for Web Performance: Metrics, Models, and Methods.* Prentice Hall, 1998.

[555] Filippo Menczer and Richard K. Belew. Adaptive information agents in distributed textual environments. In Katia P. Sycara and Michael Wooldridge, editors, *Proc. of the 2nd Int. Conference on Autonomous Agents*, pages 157–164, May 1998.

[556] A. Mendelzon, G. Mihaila, and T. Milo. Querying the World Wide Web. *International Journal on Digital Libraries*, 1(1):54–67, April 1997.

[557] Beth Meyer, Richard A. Sit, Victoria A. Spaulding, Sherry E. Mead, and Neff Walker. Age group differences in World Wide Web navigation. In Katia P. Sycara and Michael Wooldridge, editors, *Proc. of ACM Conference on Human Factors in Computing Systems*, pages 157–164, Atlanta, GA, USA, March 1997.

[558] A. Michard. Graphical presentation of Boolean expressions in a database query language: design notes and an ergonomic evaluation. *Behaviour and Information Technology*, 1(3):279–288, 1982.

[559] F. Miksa and P. Doty. Intellectual Realities and the Digital Library. In *Proc. of the 1st Annual Conference on the Theory and Practice of Digital Libraries*, College Station, TX, USA, June 1994.

[560] E. J. Miller. CNI/OCLC Metadata Workshop: Workshop on Metadata for Networked Images, September 1996. http://purl.oclc.org/metadata/image.

[561] G. A. Miller, E. B. Newman, and E. A. Friedman. Length-frequency statistics for written english. *Information and Control*, 1:370–389, 1958.

[562] J. Minker, G. A. Wilson, and B. H. Zimmerman. An evaluation of query expansion by the addition of clustered terms for a document retrieval system. *Information Storage and Retrieval*, 8(6):329–348, 1972.

[563] T. Minohara and R. Watanabe. Queries on structure in hypertext. In D. Lomet, editor, *Proc. of Foundations of Data Organization and Algorithms*, number 730 in LNCS, pages 394–411, Chicago, IL, USA, 1993. Springer-Verlag.

[564] Richard Mitchell, David Day, and Lynette Hirschman. Fishing for information on the Internet. In *Proc. of the '95 Information Visualization*, pages 105–111, Atlanta, USA, October 1995.

[565] Tom M. Mitchell. *Machine Learning.* McGraw-Hill, 1997.

[566] Mandar Mitra, Amit Singhal, and Chris Buckley. Improving automatic query expansion. In B. Croft, A. Moffat, C. J. van Rijsbergen, R. Wilkinson, and J. Zobel, editors, *Proc. of 21st Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 206–214, Melbourne, Australia, 1998.

[567] S. Miyamoto, T. Miyake, and K. Nakayama. Generation of a pseudothesaurus for information retrieval based on co-occurrences and fuzzy set operations. *IEEE Transactions on Systems and Man Cybernetics*, 13(1):62–70, 1983.

[568] S. Miyamoto and K. Nakayama. Fuzzy information retrieval based on a fuzzy pseudothesaurus. *IEEE Transactions on Systems and Man Cybernetics*, 16(2):278–282, 1986.

[569] S. Mizzaro. Relevance: The whole history. *Journal of the American Society for Information Science*, 48(9):810–832, 1997.

[570] W. E. Moen. Accessing distributed cultural heritage information. *Communications of the ACM*, 41(4):45–48, April 1998.

[571] A. Moffat. Word-based text compression. *Software Practice and Experience*, 19(2):185–198, 1989.

[572] A. Moffat and T. Bell. In situ generation of compressed inverted files. *Journal of the American Society for Information Science*, 46(7):537–550, 1995.

[573] A. Moffat and J. Zobel. Parameterized compression for sparse bitmaps. In *Proc. of the ACM SIGIR International Conference on Research and Development in Information Retrieval*, pages 274–285, Copenhagen, Denmark, 1992.

[574] Alistair Moffat and Justin Zobel. Information retrieval systems for large document collections. In Donna K. Harman, editor, *The Third Text REtrieval Conference (TREC-3)*, pages 85–94, Gaithersburg, MD, USA, 1995. Dept of Commerce, National Institute of Standards and Technology. Special Publication 500-226.

[575] E. S. Moura, G. Navarro, and N. Ziviani. Indexing compressed text. In R. Baeza-Yates, editor, *Proc. of the 4th South American Workshop on String Processing*, volume 8, pages 95–111, Valparaíso, Chile, 1997. Carleton University Press International Informatics Series.

[576] E. S. Moura, G. Navarro, N. Ziviani, and R. Baeza-Yates. Direct pattern matching on compressed text. In *Proc. of the 5th Symposium on String Processing and Information Retrieval*, pages 90–95, Santa Cruz, Bolivia, September 1998.

[577] E. S. Moura, G. Navarro, N. Ziviani, and R. Baeza-Yates. Fast searching on compressed text allowing errors. In *Proc. of the ACM-SIGIR International Conference on Research and Development in Information Retrieval*, pages 298–306, Melbourne, Australia, August 1998.

[578] S. Mukherjea and J. D. Foley. Visualizing the World Wide Web with the Navigational View Builder. *Computer Networks and ISDN Systems*, 27:1075–1087, 1995.

[579] S. Mukherjea, K. Hirata, and Y. Hara. Towards a multimedia World-Wide Web information retrieval engine. In *6th Int. WWW Conference*, Santa Clara, CA, USA, April 1997.

[580] Kevin Mullet and Darell Sano. *Designing Visual Interfaces: Communication-Oriented Techniques.* SunSoft Press, 1995.

[581] Brad Myers. *Creating User Interfaces by Demonstration.* Academic Press, New York, 1988.

[582] Brad Myers. A taxonomy of window manager user interfaces. *IEEE Computer Graphics and Applications*, pages 65–84, Sept 1988.

[583] G. Myers. A fast bit-vector algorithm for approximate pattern matching based on dynamic programming. In M. Farach-Colton, editor, *Proc. of Combinatorial Pattern Matching*, number 1448 in LNCS, pages 1–13, Rutgers, USA, 1998. Springer-Verlag.

[584] Bonnie A. Nardi. *A Small Matter of Programming: Perspectives on End User Computing*. MIT Press, 1993.

[585] A. Nation. Visualizing Websites using a hierarchical table of contents browser: WebTOC. In *Proc. of the 3rd Conference on Human Factors and the Web*, Denver, CO, 1997.

[586] National Humanities Alliance, Basic Principles for Managing Intellectual Property in the Digital Environment, March 1997. http://www-ninch.cni.org/ISSUES/COPYRIGHT/-PRINCIPLES/NHA_Complete.html.

[587] *Nation's Information Infrastructure Steering Committee, Computer Science, More than screen deep: toward every-citizen interfaces to the nation's information infrastructure.* National Academy Press, Washington, DC, USA, 1997.

[588] G. Navarro. *Approximate Text Searching*. PhD thesis, Dept. of Computer Science, Univ. of Chile, December 1998.

[589] G. Navarro and R. Baeza-Yates. A language for queries on structure and contents of textual databases. In *Proc. of the 18th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 93–101, Seattle, USA, July 1995.

[590] G. Navarro and R. Baeza-Yates. Proximal nodes: A model to query document databases by content and structure. *ACM Transactions on Office and Information Systems*, 15(4):401–435, 1997.

[591] G. Navarro, J. Kitajima, B. Ribeiro-Neto, and N. Ziviani. Distributed generation of suffix arrays. In A. Apostolico and J. Hein, editors, *Proc. of Combinatorial Pattern Matching*, number 1264 in LNCS, pages 102–115, Aarhus, Denmark, 1997. Springer-Verlag.

[592] G. Navarro and M. Raffinot. A bit-parallel approach to suffix automata: fast extended string matching. In M. Farach-Colton, editor, *Proc. of Combinatorial Pattern Matching*, number 1448 in LNCS, pages 14–33, Rutgers, USA, 1998. Springer-Verlag.

[593] G. Navarro and M. Raffinot. A general practical approach to pattern matching over Ziv-Lempel compressed text. Technical Report TR/DCC-98-12, Dept. of Computer Science, Univ. of Chile, 1998. ftp://ftp.dcc.uchile.cl/pub/users/gnavarro/lzsrch.ps.gz.

[594] NDLTD. Networked Digital Library of Theses and Dissertations, June 1998. http://www.ndltd.org/.

[595] J. Nesbit. The accuracy of approximate string matching algorithms. *Journal of Computer-Based Instruction*, 13(3):80–83, 1986.

[596] Netcraft Web Server Survey. http://www.netcraft.com/Survey/, 1998.

[597] NetSizer: Main Page. http://www.netsizer.com, 1998.

[598] Network Development and MARC Standards Office, Dublin Core/MARC/GILS Crosswalk, July 1997. http://www.loc.gov/marc/dccross.html.

[599] Network Wizards: Internet Domain Survey. http://www.nw.com, 1998.

[600] M. Lynne Neufeld and Martha Cornog, Database history: from dinosaurs to compact discs. *Journal of the American Society for Information Science*, 37(4):183–190, 1986.

[601] New Zealand DL Group, The New Zealand Digital Library Project. http://www.nzdl.org/.

[602] D. Ngu and X. Wu. SiteHelper: a localized agent that helps incremental exploration of the World Wide Web. In *6th. Int. WWW Conference*, Santa Clara, CA, USA, April 1997.

[603] Wayne Niblack, Ron Barber, Will Equitz, Myron Flickner, Eduardo Glasman, Dragutin Petkovic, Peter Yanker, Christos Faloutsos, and Gabriel Taubin. The QBIC project: Querying images by content using color, texture and shape. In *Int. Symposium on Electronic Imaging: Science and Technology, Conf. 1908, Storage and Retrieval for Image and Video Databases*, February 1993.

[604] Jakob Nielsen. *Hypertext and Hypermedia*. Academic Press, 1990.

[605] Jakob Nielsen. *Usability Engineering*. Academic Press, 1993.

[606] NISO Press, Bethesda, MD. *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification (ANSI/NISO Z39.50-1995)*, 1995. Accessible at http://lcweb.loc.gov/z3950/agency/.

[607] D. A. Norman. *The Psychology of Everyday Things*. Basic Books, New York, 1988.

[608] Northern Light: Main page. http://www.northernlight.com, 1997.

[609] G. Notess. Search Engines Showdown. http://imt.net/~notess/search/, 1998.

[610] L. Nowell. *Graphical Encoding for Information Visualization: Using Icon Color, Shape, and Size to Convey Nominal and Quantitative Data*. PhD Thesis, Virginia Polytechnic and State University, Department of Computer Science, 1997.

[611] P. J. Nuernberg, R. Furuta, J. J. Leggett, C. C. Marshall, and F. M. Shipman III. Digital Libraries: Issues and Architectures. In *Proc. of the 2nd Annual Conference on the Theory and Practice of Digital Libraries*, Austin, Texas, 1995.

[612] D. W. Oard. Serving Users in Many Languages: Cross-Language Information Retrieval for Digital Libraries. *D-Lib Magazine*, December 1997. http://www.dlib.org/dlib/december97/oard/12oard.html.

[613] D. W. Oard and B. J. Dorr. A survey of multilingual text retrieval. Technical Report UMIACS-TR-96-19, University of Maryland, Institute for Advanced Computer Studies, 1996.

[614] Vicki L. O'Day and Robin Jeffries. Orienteering in an information landscape: how information seekers get from here to there. In *Proc. of the INTERCHI '93*, Amsterdam, Netherlands, April 1993. IOS Press.

[615] R. N. Oddy. Information retrieval through man-machine dialogue. *Journal of Documentation*, 33:1–14, 1977.

[616] Y. Ogawa, T. Morita, and K. Kobayashi. A fuzzy document retrieval system using the keyword connection matrix and a learning method. *Fuzzy Sets and Systems*, 39:163–179, 1991.

[617] Kenton O'Hara and Abigail Sellen. A comparison of reading paper and on-line documents. In *Proc. of ACM Conference on Human Factors in Computing Systems*, Atlanta, GA, USA, March 1997. http://www.acm.org/sigchi/chi97/proceedings/paper/koh.htm.

[618] K. Olsen, R. Korfhage, K. Sochats, M. Spring, and J. Williams. Visualization of a document collection with implicit and explicit links: The Vibe system. *Scandinavian Journal of Information Systems*, 5, 1993.

[619] Ed O'Neill, Brian Lavoie, and Pat McClain. OCLC Web characterization project (position paper). In *Web Characterization Workshop*, Boston, USA, Nov 1998. http://www.w3.org/1998/11/05/WC-workshop/Papers/oneill.htm.

[620] Online Computer Library Center, Inc. (OCLC). Metadata Workshop II, April 1996. http://www.oclc.org:5046/oclc/research/conferences/metadata2/ .

[621] E. Oomoto and K. Tanaka. OVID: Design and implementation of a video-object database system. *IEEE Trans. on Knowledge and Data Engineering.*, 5:629–641, 1993.

[622] Alan Victor Oppenheim and Ronald W. Schafer. *Digital Signal Processing.* Prentice Hall, Englewood Cliffs, NJ, 1975.

[623] J. A. Orenstein. A Comparison of Spatial Query Processing Techniques for Native and Parameter Spaces. In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, pages 343–352, Atlantic City, New Jersey, 1988.

[624] A. Paepcke, C.-C. K. Chang, H. García-Molina, and T. Winograd. Interoperability for Digital Libraries Worldwide. *Communications of the ACM*, 41(4):33–43, April 1998.

[625] A. Paepcke, S. B. Cousins, H. García-Molina, S. W. Hassan, and S. P. Ketchpel. Using distributed objects for digital library interoperability. *IEEE Computer*, 29(5):61–68, May 1996.

[626] C. D. Paice and P. A. Jones. The identification of important concepts in highly structured technical papers. In *Proc. of the 16th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 69–78, Pittsburgh, USA, 1993.

[627] George Panagopoulos and Christos Faloutsos. Bit-sliced signature files for very large text databases on a parallel machine architecture. In *Proc. 4th Int. Conf. on Extending Database Technology (EDBT)*, number 779 in LNCS, pages 379–392, London, 1994. Springer-Verlag.

[628] Apostolos Papadopoulos and Yannis Manolopoulos. Performance of nearest neighbor queries in R-trees. In Foto N. Afrati and Phokion Kolaitis, editors, *Proc. of 6th Int. Conf. on Database Theory*, number 1186 in LNCS, pages 394–408, Delphi, Greece, Jan 1997.

[629] Michael Pazzani, Daniel Billsus, and Jack Muramatsu. Syskill & Webert: Identifying interesting Web sites. In *Proc. of the Thirteenth Annual National Conference on Artificial Intelligence*, pages 54–61, Portland, OR, USA, August 1996.

[630] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Inc., 1988.

[631] Gert Schmeltz Pedersen. A browser for bibliographic information retrieval, based on an application of lattice theory. In *Proc. of the 16th Annual Int. ACM SIGIR Conference*, pages 270–279, Pittsburgh, PA, 1993.

[632] Theresa Pepin, Jeff Barry, and W. David Penniman. Automated system marketplace: the competitive edge: expanded access drives vendors. *Library Journal*, 122(6):47–56, 1997.

[633] C. Peters and E. Picchi. across languages, across cultures: Issues in multilinguality and digital libraries. *D-Lib Magazine*, May 1997. http://www.dlib.org/dlib/may97/peters/05peters.html.

[634] P. E. Peters. Digital libraries are much more than digitized collections. *EDUCOM Review*, 30(4), 1995.

[635] Euripides G. M. Petrakis and Christos Faloutsos. Similarity searching in medical image databases. *IEEE Trans. on Knowledge and Data Engineering*, 9(3):435–447, May 1997.

[636] Lori Phelps. Active documentation: Wizards as a medium for meeting user needs. In *Proc. of the ACM 15th International Conference on Systems Documentation*, pages 207–210, 1997.

[637] D. Pimienta. Languages, culture, and Internet (in French). http://funredes.org/, March 1998.

[638] Peter Pirolli and Stuart Card. Information foraging models of browsers for very large document spaces. In *Advanced Visual Interfaces*, L'Aquila, Italy, May 1998.

[639] Peter Pirolli, James Pitkow, and Ramana Rao. Silk from a sow's ear: Extracting usable structures from the Web. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 118–125, Zurich, Switzerland, May 1996. ACM Press.

[640] Peter Pirolli, Patricia Schank, Marti A. Hearst, and Christine Diehl. Scatter/gather browsing communicates the topic structure of a very large text collection. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 213–220, Zurich, Switzerland, May 1996.

[641] J. E. Pitkow. Summary of WWW characterizations. In *7th WWW Conference*, Brisbane, Australia, 1998.

[642] J. E. Pitkow and K. A. Bharat. WebViz: A tools for World Wide Web access log analysis. In *Proc. of the 1st Int. WWW Conference*, Geneva, Switzerland, May 1994. http://www1.cern.ch/PapersWWW94/pitkow-webvis.ps.

[643] Catherine Plaisant, Tom Bruns, Ben Shneiderman, and Khoa Doan. Query previews in networked information systems: the case of EOSDIS. In *Proc. of ACM Conference on Human Factors in Computing Systems*, volume 2 of *Formal Video Program*, pages 202–203, Atlanta, GA, USA, March 1997.

[644] Catherine Plaisant, David Carr, and Ben Shneiderman. Image-browser taxonomy and guidelines for designers. *IEEE Software*, 12(2):21–32, 1995.

[645] Christine A. Pogue and Peter Willet. Use of text signatures for document retrieval in a highly parallel environment. *Parallel Computing*, 4:259–268, 1987.

[646] Steven Pollitt. Interactive information retrieval based on faceted classification using views. In *Proc. of the 6th Int. Study Conference on Classification (FID/CR)*, University College, London, June 1997.

[647] A. Pollock and A. Hockley. What's wrong with Internet searching. *D-Lib Magazine*, March 1997.

[648] M. F. Porter. An algorithm for suffix stripping. In K. Sparck Jones and P. Willet, editors, *Readings in Information Retrieval*, pages 313–316. Morgan Kaufmann Publishers, Inc., 1997.

[649] J. Powell and E. Fox. Multilingual federated searching across heterogeneous collections. *D-Lib Magazine*, September 1998.

[650] Wanda Pratt. Dynamic organization of search results using the UMLS. In *American Medical Informatics Association Fall Symposium*, Nashville, TN, USA, October 1997.

[651] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.

[652] G. Pringle, L. Allison, and D. L. Dowe. What is a tall poppy among Web pages? In *7th WWW Conference*, Brisbane, Australia, April 1998.

[653] Teresa Pritchard-Schoch. Comparing natural language retrieval: WIN & Freestyle. *Online*, 19(4):83–87, 1995.

[654] The PURL Team, Persistent Uniform Resource Locator (PURL). http://purl.oclc.org/.

[655] Yonggang Qiu and H. P. Frei. Concept based query expansion. In *Proc. of the 16th ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 160–169, Pittsburgh, PA, USA, 1993.

[656] D. Quass, A. Rajaraman, Y. Sagiv, J. Ullman, and J. Widom. Querying semistructured heterogeneous information. In *Deductive and Object-Oriented Databases, Proc. of the DOOD '95 Conference*, pages 319–344, Singapore, December 1995. Springer-Verlag.

[657] F. Rabitti and P. Savino. Retrieval of multimedia document by imprecise query specification. In F. Bancilhon, C. Thanos, and D. Tsichritzis, editors, *Proc. of the Int. Conference on Extended Database Technologies*, number 416 in LNCS, pages 203–218, Venice, Italy, 1990.

[658] T. Radecki. Mathematical model of information retrieval system based on the concept of fuzzy thesaurus. *Information Processing & Management*, 12:313–318, 1976.

[659] T. Radecki. Mathematical model of time-effective information retrieval system based on the theory of fuzzy sets. *Information Processing & Management*, 13:109–116, 1977.

[660] T. Radecki. Fuzzy set theoretical approach to document retrieval. *Information Processing & Management*, 15:247–259, 1979.

[661] T. Radecki. On the inclusiveness of information retrieval systems with documents indexed by weighted descriptors. *Fuzzy Sets and Systems*, 5:159–176, 1981.

[662] Tadeusz Radecki. Trends in research on information retrieval—the potential for improvements in conventional Boolean retrieval systems. *Information Processing & Management*, 24:219–227, 1988.

[663] V. V. Raghavan, P. Bollmann, and G. S. Jung. Retrieval system evaluation using recall and precision: Problems and answers. In *Proc. of the 12th ACM SIGIR Conference*, pages 59–68, Cambridge, USA, June 1989.

[664] V. V. Raghavan, G. S. Jung, and P. Bollmann. A critical investigation of recall and precision as measures of retrieval system performance. *ACM Transactions on Office and Information Systems*, 7(3):205–229, 1989.

[665] V. V. Raghavan and S. K. M. Wong. A critical analysis of vector space model for information retrieval. *Journal of the American Society for Information Sciences*, 37(5):279–287, 1986.

[666] R. Rao and S. K. Card. The Table Lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 318–322, Boston, MA, USA, April 1994.

[667] R. Rao, S. K. Card, H. D. Jellinek, J. D. Mackinlay, and G. G. Robertson. The Information Grid: A framework for building information retrieval and retrieval-centered applications. In *Proc. of the ACM Symposium on User Interface Software and Technology*, Monterey, CA, USA, Nov 1992.

[668] E. Rasmussen. Clustering algorithms. In W. B. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures & Algorithms*, pages 419–442. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[669] David F. Redmiles. Reducing the variability of programmers' performance through explained examples. In *Proc. of the ACM Conference on Human Factors in Computing Systems*, pages 67–73, 1993.

[670] A. Renear. The digital library research agenda: What's missing — and how humanities textbase projects can help. *D-Lib Magazine*, July/August 1997. http://www.dlib.org/dlib/july97/07renear.html.

[671] Earl Rennison. Galaxy of news: An approach to visualizing and understanding expansive news landscapes. In *Proc. of UIST'94, ACM Symposium on User Interface Software and Technology*, pages 3–12, New York, 1994.

[672] Paul Resnick and Hal Varian. Introduction: Special issue on collaborative filtering. *Communications of the ACM*, 40(3):56–58, March 1997.

[673] B. Ribeiro-Neto and R. Barbosa. Query performance in tightly coupled distributed digital libraries. In *Proc. of the 3rd Annual Int. ACM Conference on Digital Libraries*, Pittsburgh, PA, USA, 1998.

[674] Berthier A. Ribeiro-Neto and Richard Muntz. A belief network model for IR. In *Proc. of the 19th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 253–260, Zurich, Switzerland, 1996.

[675] J. Rissanen and G. G. Langdon. Arithmetic coding. *IBM Journal of Research and Development*, 23:149–162, 1979.

[676] George C. Robertson, Stuart K. Card, and Jock D. Mackinlay. Information visualization using 3D interactive animation. *Communications of the ACM*, 36(4):56–71, 1993.

[677] S. E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Sciences*, 27(3):129–146, 1976.

[678] J. J. Rocchio. Relevance feedback in information retrieval. In G. Salton, editor, *The SMART Retrieval System — Experiments in Automatic Document Processing*. Prentice Hall Inc., Englewood Cliffs, NJ, 1971.

[679] Peter Roget. *Roget's II The New Thesaurus*. Houghton Mifflin Company, Boston, USA, 1988.

[680] M. Roscheisen, C. Mogensen, and T. Winograd. Interaction design for shared World-Wide Web annotations. Stanford Digital Library Project Working Paper, February 1995. http://walrus.stanford.edu/diglib/pub/reports/brio-chi95.html .

[681] David E. Rose, Richard Mander, Tim Oren, Dulce B. Ponceleón, Gitta Salomon, and Yin Yin Wong. Content awareness in a file system interface: Implementing the 'pile' metaphor for organizing information. In *Proc. of the 16th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 260–269, 1993.

[682] Louis Rosenfeld and Peter Morville. *Information Architecture for the World Wide Web*. O'Reilly & Associates, 1998.

[683] A. Ross. Library functions, scholarly communication, and the foundation of the digital library: Laying claim to the control zone. *The Library Quarterly*, 66:239–265, July 1996.

[684] Mary Beth Rosson, John M. Carroll, and Rachel K. E. Bellamy. Smalltalk Scaffolding: A case study of minimalist instruction. In *Proc. of the ACM Conference on Human Factors in Computing Systems*, pages 423–429, 1990.

[685] Steven F. Roth, Mei C. Chuah, Stephan Kerpedjiev, John A. Kolojejchick, and Peter Lucas. Towards an information visualization workspace: Combining multiple means of expression. *Human-Computer Interaction*, 12(1-2):131–185, 1997.

[686] Nick Roussopoulos, Steve Kelley, and F. Vincent. Nearest neighbor queries. In *Proc. of ACM-SIGMOD*, pages 71–79, San Jose, CA, May 1995.

[687] N. Roussopoulus, C. Faloutsos, and T. Sellis. An efficient pictorial database system for PSQL. *IEEE Transactions on Software Engineering*, 14(5):639–650, May 1988.

[688] Jennifer Rowley. The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research. *Journal of Information Science*, 20(2):108–119, 1994.

[689] Mary Beth Ruskai, Gregory Beylkin, Ronald Coifman, Ingrid Daubechies, Stephane Mallat, Yves Meyer, and Louise Raphael. *Wavelets and their Applications*. Jones and Bartlett Publishers, Boston, MA, 1992.

[690] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. The cost structure of sensemaking. In *Proc. of ACM Conference on Human Factors in Computing Systems*, pages 269–276, 1993.

[691] W. M. Sachs. An approach to associative retrieval through the theory of fuzzy sets. *Journal of the American Society for Information Sciences*, pages 85–87, 1976.

[692] M. Sahami, S. Yusufali, and M. Q. W. Baldonado. SONIA: A service for organizing networked information autonomously. In *Proc. of the 3rd Annual Conference on Digital Libraries*, pages 200–209, New York, USA, 1998.

[693] A. Salminen and F. Tompa. PAT expressions: an algebra for text search. Technical Report OED-92-02, UW Centre for the New Oxford English Dictionary, July 1992.

[694] G. Salton. Associative document retrieval techniques using bibliographic information. *Journal of the ACM*, 10(4):440–457, October 1963.

[695] G. Salton. *The SMART Retrieval System — Experiments in Automatic Document Processing*. Prentice Hall Inc., Englewood Cliffs, NJ, 1971.

[696] G. Salton and C. Buckley. Term-weighting approaches in automatic retrieval. *Information Processing & Management*, 24(5):513–523, 1988.

[697] G. Salton and M. E. Lesk. Computer evaluation of indexing and text processing. *Journal of the ACM*, 15(1):8–36, January 1968.

[698] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill Book Co., New York, 1983.

[699] G. Salton, C. S. Yang, and C. T. Yu. A theory of term importance in automatic text analysis. *Journal of the American Society for Information Sciences*, 26(1):33–44, 1975.

[700] Gerard Salton. *Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, Reading, MA, 1989.

[701] Gerard Salton and Chris Buckley. Parallel text search methods. *Communications of the ACM*, 31(2):202–215, February 1988.

[702] Gerard Salton and Chris Buckley. Improving retrieval performance by relevance feedback. *J. of the American Society for Information Science*, 41(4):288–297, 1990.

[703] Gerard Salton, Edward A. Fox, and Harry Wu. Extended Boolean information retrieval. *Communications of the ACM*, 26(11):1022–1036, November 1983.

[704] Gerard Salton and C. S. Yang. On the specification of term values in automatic indexing. *Journal of Documentation*, 29:351–372, 1973.

[705] P. Samuelson. Copyright and digital libraries. In *Proc. of the 2nd ACM International Conference on Digital Libraries*, pages 113–114, Philadelphia, PA, USA, 1997.

[706] David Sankoff and Joseph B. Kruskal. *Time warps, string edits and macromolecules: The theory and practice of sequence comparisons*. Addison-Wesley, Reading, MA, 1983.

[707] Tefko Saracevic, Paul Kantor, Alice Y. Chamis, and Donna Trivison. A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science*, 39(3):161–176, 1988.

[708] S. Sato and T. Kanade. Name-it: Association of Face and Name in Video. Technical Report CMU-CS-96-205, CMU School of Computer Science, 1996.

[709] L. Schamber. What is a document? Rethinking the concept in uneasy times. *Journal of the American Society for Information Science*, 47:669–671, September 1996.

[710] B. Schatz and H. Chen. Building large-scale digital libraries: Guest editors' introduction to theme issue on the US Digital Library Initiative. *IEEE Computer*, 29(5):22–26, May 1996. http://computer.org/computer/dli/.

[711] B. R. Schatz. Information retrieval in digital libraries: Bringing search to the Net. *Science*, 275:327–335, January 1997.

[712] Manfred Schroeder. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. W. H. Freeman and Company, New York, 1991.

[713] E. S. Schwartz and B. Kallick. Generating a canonical prefix encoding. *Communications of the ACM*, 7:166–169, 1964.

[714] S. Sclaroff, L. Taycher, and M. La Cascia. ImageRover: A content-based image browser for the World Wide Web. In *Proc. IEEE Workshop on Content-based Access of Image and Video Libraries*, San Juan, Puerto Rico, June 1997.

[715] E. Selberg and O. Etzioni. Multi-service search and comparison using the MetaCrawler. *4th Int. WWW Conference*, December 1995.

[716] P. Sellers. The theory and computation of evolutionary distances: pattern recognition. *Journal of Algorithms*, 1:359–373, 1980.

[717] T. Sellis, N. Roussopoulos, and C. Faloutsos. The R^+-tree: A dynamic index for multi-dimensional objects. In *Proc. of the 13th Int. Conference on Very Large Data Bases*, pages 507–518, Brighton, England, 1987.

[718] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:398–403, 1948.

[719] Jacob Shapiro, Vladimir G. Voiskunskii, and Valery J. Frants. *Automated Information Retrieval : Theory and Text-Only Methods*. Academic Press, 1997.

[720] Upendra Shardanand and Pattie Maes. Social information filtering: Algorithms for automating Word of Mouth. In *Proc. of ACM Conference on Human Factors in Computing Systems*, pages 210–217, Denver, CO, USA, 1995.

[721] W. M. Shaw, J. B. Wood, R. E. Wood, and H. R. Tibbo. The cystic fibrosis database: Content and research opportunities. *Library and Information Science Research*, 13:347–366, 1991.

[722] Frank M. Shipman, III, Catherine C. Marshall, and Thomas P. Moran. Finding and using implicit structure in human-organized spatial layouts of information. In *Proc. of ACM Conference on Human Factors in Computing Systems*, volume 1, pages 346–353, Denver, CO, USA, 1995.

[723] N. Shivakumar and H. García-Molina. Finding near-replicas of documents on the Web. In *Workshop on Web Databases*, Valencia, Spain, March 1998.

[724] Ben Shneiderman. The eyes have it: A task by data type taxonomy. In *Proc. of IEEE Symp. Visual Languages 96*, pages 336–343, Boulder, CO, USA, 1996.

[725] Ben Shneiderman. *Designing the User Interface: Strategies for Effective Human-computer Interaction*. Addison-Wesley, Reading, MA, 1997.

[726] Ben Shneiderman, Donald Byrd, and W. Bruce Croft. Sorting out searching: A user-interface framework for text searches. *Communications of the ACM*, 41(4):95–98, 1998.

[727] Ben Shneiderman and Greg Kearsley. *Hypertext Hands-On! An Introduction to a New Way of Organizing and Accessing Information*. Addison-Wesley Publishing Co., Reading, MA, 1989.

[728] C. Silverstein, M. Henzinger, J. Marais, and M. Moricz. Analysis of a very large alta vista query log. Technical Report 1998-014, COMPAQ Systems Research Center, Palo Alto, CA, USA, 1998.

[729] Singapore advanced research and education network (singaren). http://www.singaren.net.sg/.

[730] H. Small. Co-citation in the scientific literature: A new measure of relationship between two documents. *Journal of the American Society for Information Science*, 24(4):265–269, 1973.

[731] H. Small. The relationship of information science to the social sciences: A co-citation analysis. *Information Processing & Management*, 17(1):39–50, 1981.

[732] H. Small and M. E. D. Koenig. Journal clustering using a bibliographic coupling method. *Information Processing & Management*, 13(5):277–288, 1977.

[733] D. C. Smith and S.-F. Chang. Searching for images and videos on the World-Wide Web. *IEEE Multimedia Magazine*, 4(3):12–20, 1997.

[734] T. R. Smith. The Meta-information environment of digital libraries. *D-Lib Magazine*, July/August 1996. http://www.dlib.org/dlib/july96/new/07smith.html.

[735] Dagobert Soergel. *Indexing Languages and Thesauri: Construction and Maintenance.* Melville Publishing Co., Los Angeles, CA, 1974.

[736] Ellen Spertus. ParaSite: Mining structural information on the Web. In *6th Int. WWW Conference*, Santa Clara, CA, USA, April 1997.

[737] M. Spiliopoulou and L. Faulstich. WUM — A tool for WWW utilization analysis. In *Workshop on Web Databases*, pages 109–115, Valencia, Spain, March 1998.

[738] Anselm Spoerri. InfoCrystal: A visual tool for information retrieval & management. In *Proc. of Information Knowledge and Management '93*, pages 11–20, Washington, DC, Nov 1993.

[739] P. Srinivasdan. Thesaurus construction. In W. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures & Algorithms*, pages 161–218. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[740] Craig Stanfill. Partitioned posting files: A parallel inverted file structure for information retrieval. In *Proc. 13th Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 413–428, Brussels, Belgium, 1990.

[741] Craig Stanfill. Parallel information retrieval algorithms. In William B. Frakes and Ricardo Baeza-Yates, editors, *Information Retrieval Data Structures & Algorithms*, pages 459–497. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[742] Craig Stanfill and Brewster Kahle. Parallel free-text search on the Connection Machine system. *Communications of the ACM*, 29(12):1229–1239, December 1986.

[743] Craig Stanfill and Brewster Kahle. Toward memory-based reasoning. *Communications of the ACM*, 29(12):1213–1228, 1986.

[744] Craig Stanfill, Robert Thau, and David Waltz. A parallel indexed algorithm for information retrieval. In *Proc. 12th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 88–97, Cambridge, USA, June 1989.

[745] Stanford DLI Team, Stanford University Digital Libraries Project. http://www-diglib.stanford.edu/diglib/.

[746] M. Stefik. Trusted systems. *Scientific American*, March 1997. http://www.sciam.com/-0397issue/0397stefik.html.

[747] J. A. Storer. *Data Compression: Methods and Theory.* Computer Science Press, 1988.

[748] Tomek Strzalkowski, editor. *Natural Language Information Retrieval.* Kluwer Academic Publishers, 1999. To appear.

[749] D. Sullivan. Search Engine Watch: Main page. http://www.searchenginewatch.com, 1997.

[750] K. Summers. Toward a taxonomy of logical document structures. In *DAGS95: Electronic Publishing and the Information Superhighway, May 30–June 2, 1995*, 1995. http://www.cs.dartmouth.edu/~samr/DAGS95/Papers/summers.html .

[751] D. Sunday. A very fast substring search algorithm. *Communications of the ACM*, 33(8):132–142, August 1990.

[752] Elaine Svenonius. Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science*, 37(5):331–340, 1986.

[753] R. Swick. Resource Description Framework (RDF), June 1998. http://www.w3.org/RDF.

[754] J. Tague-Sutcliffe. Measuring the informativeness of a retrieval process. In *Proc. of the 15th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 23–36, Copenhagen, Denmark, 1992.

[755] V. A. Tahani. A fuzzy model of document retrieval systems. *Information Processing & Management*, 12:177–187, 1976.

[756] TEI. A gentle introduction to SGML. Technical report, Text Encoding Initiative, 1996. http://www.sil.org/sgml/gentle.html.

[757] Carol Tenopir and Jeff Barry. Database marketplace: data dealers face stormy weather. *Library Journal*, 123(9):38–46, 1998.

[758] Carol Tenopir and Pamela Cahn. TARGET & FREESTYLE: DIALOG and Mead join the relevance ranks. *Online*, 18(3):31–47, 1994.

[759] C. Thanos, editor. *Multimedia Office Filing: The MULTOS Approach*. Elsevier Science Publishers B.V. (North Holland), 1990.

[760] P. Thompson, H. Turtle, B. Yang, and J. Flood. TREC-3 ad-hoc retrieval and routing experiments using the WIN system. In *Proc. of the Text REtrieval Conference*, pages 211–218, Gaithersburg, MD, USA, 1995.

[761] R. H. Thompson and B. W. Croft. Support for browsing in an intelligent text retrieval system. *International Journal of Man-Machine Studies*, 30(6):639–668, 1989.

[762] Anthony Tomasic and Hector García-Molina. Caching and database scaling in distributed shared-nothing information retrieval systems. In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, pages 129–138, Washington, DC, USA, May 1993.

[763] Anthony Tomasic and Hector García-Molina. Performance of inverted indices in shared-nothing distributed text document information retrieval systems. In *Proc. of the 2nd Int. Conf. on Parallel and Distributed Information Systems*, pages 8–17, San Diego, USA, 1993.

[764] Anthony Tomasic and Hector García-Molina. Performance issues in distributed shared-nothing information retrieval systems. *Information Processing & Management*, 32(6):647–665, 1996.

[765] Anthony Tomasic, Luis Gravano, Calvin Lue, Peter Schwarz, and Laura Haas. Data structures for efficient broker implementation. *ACM Transactions on Information Systems*, 15(3):223–253, 1997.

[766] A. Tombros and M. Sanderson. Advantages of query biased summaries in Information Retrieval. In B. Croft, A. Moffat, C. J. van Rijsbergen, R. Wilkinson, and J. Zobel, editors, *Proc. of the 21st ACM SIGIR Conference on Reseach and Development in Information Retrieval*, pages 2–10, Melbourne, Australia, 1998.

[767] Geoffrey Towell, Ellen M. Voorhees, Narendra K. Gupta, and Ben Johnson-Laird. Learning collection fusion strategies for information retrieval. In *Proc. of the 12th Int. Conference on Machine Learning*, pages 540–548. Morgan Kaufmann, 1995.

[768] The TREC NIST site, 1998. http://trec.nist.gov.

[769] Edward Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Chelshire, CT, 1983.

[770] Edward Tufte. *Envisioning Information*. Graphics Press, Chelshire, CT, 1990.

[771] H. Turtle and W. B. Croft. Evaluation of an inference network-based retrieval model. *ACM Transactions on Information Systems*, 9(3):187–222, July 1991.

[772] Howard Turtle and W. Bruce Croft. Inference networks for document retrieval. In *Proc. of the 13th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1–24, Brussels, Belgium, 1990.

[773] L. A. Tweedie, R. Spence, D. Williams, and R. Bhogal. The attribute explorer. In *Proc. of the Video Track of the ACM Conference on Human Factors in Computing Systems*, pages 435–436, Boston, MA, USA, April 1994.

[774] UC Berkeley Digital Library Project, About MVD version 1.0alpha3. http://elib.cs.berkeley.edu/java/help/About.html.

[775] UC Berkeley DLI Team, UC Berkeley Digital Library Project. http://elib.cs.berkeley.edu/.

[776] UC Santa Barbara DLI Team, Alexandria Digital Library. http://alexandria.sdc.ucsb.edu/.

[777] UIUC DLI Team, University of Illinois at Urbana-Champaign Digital Libraries. http://dli.grainger.uiuc.edu/default.htm.

[778] The UK Office for Library and Information Networking, Electronic Libraries Programme, eLib, March 1998. http://www.ukoln.ac.uk/services/elib/.

[779] E. Ukkonen. Approximate string matching over suffix trees. In A. Apostolico, M. Crochemore, Z. Galil, and U. Manber, editors, *Proc. of Combinatorial Pattern Matching*, number 684 in LNCS, pages 228–242, Padova, Italy, 1993. Springer-Verlag.

[780] E. Ukkonen. Constructing suffix trees on-line in linear time. *Algorithmica*, 14(3):249-260, Sep 1995.

[781] Esko Ukkonen. Finding approximate patterns in strings. *Journal of Algorithms*, 6:132–137, 1985.

[782] G. M. Underwood, P. P. Maglio, and R. Barrett. User-centered push for timely information delivery. In *7th WWW Conference*, Brisbane, Australia, 1998.

[783] Unicode Consortium, Unicode. http://www.unicode.org/.

[784] University of Michigan DLI Team, Univ. of Michigan Digital Library Project. http://-www.si.umich.edu/UMDL/.

[785] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, 1979.

[786] J. Verhoeff, W. Goffmann, and Jack Belzer. Inefficiency of the use of Boolean functions for information retrieval systems. *Communications of the ACM*, 4(12):557–558, 594, December 1961.

[787] Charles L. Viles and James C. French. Dissemination of collection wide information in a distributed information retrieval system. In *Proc. 18th Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 12–20, Seattle, WA, USA, July 1995.

[788] S. Vinoski. CORBA: Integrating diverse applications within distributed heterogeneous environments. *IEEE Communications Magazine*, 14(2), February 1997.

[789] J. S. Vitter. Algorithm 673: Dynamic Huffman coding. *ACM Transactions on Mathematical Software*, 15:158–167, 1989.

[790] A. N. Vo and A. Moffat. Compressed inverted files with reduced decoding overhead. In *Proc. of the ACM-SIGIR Int. Conference on Research and Development in Information Retrieval*, pages 290–297, Melbourne, Australia, 1998.

[791] E. M. Voorhees and R. M. Tong. Multiple search engines in database merging. In *Proc. of the 2nd ACM International Conference on Digital Libraries*, pages 93-102, Philadelphia, PA, USA, 1997.

[792] Ellen M. Voorhees, Narendra K. Gupta, and Ben Johnson-Laird. The collection fusion problem. In Donna K. Harman, editor, *The Third Text REtrieval Conference (TREC-3)*, pages 95–104, Gaithersburg, MD, USA, 1995. Dept. of Commerce, National Institute of Standards and Technology. Special Publication 500-226.

[793] E. M. Voorhees. *The Effectiveness and Efficiency of Agglomerative Hierarchic Clustering in Document Retrieval*. PhD thesis, Cornell University, 1986.

[794] E. M. Voorhees and D. K. Harman. Overview of the 6th text retrieval conference (TREC-6). In E. M. Voorhees and D. K. Harman, editors, *Proc. of the Sixth Text REtrieval Conference*, pages 1–24. NIST Special Publication, Gaithersburg, MD, USA, 1997.

[795] W3C. Extensible markup language (XML) 1.0. Technical report, WWW Consortium (W3C), 1998. http://www.w3.org/TR/1998/REC-xml-19980210.

[796] W3C. HTML 4.0 specification. Technical report, WWW Consortium (W3C), 1998. http://www.w3.org/TR/1998/REC-html40-19980424/.

[797] W3C. Web characterization activity, 1998. http://www.w3.org/WCA/.

[798] W3C. XML linking language (XLink). Technical report, WWW Consortium (W3C), 1998. http://www.w3.org/TR/1998/WD-xlink-19980303.

[799] W3C. XSL requirements summary. Technical report, WWW Consortium (W3C), 1998. http://www.w3.org/TR/1998/WD-XSLReq-19980511.

[800] Howard D. Wactlar, Takeo Kanade, Michael A. Smith, and Scott M. Stevens. Intelligent access to digital video: Informedia project. *IEEE Computer*, 29(5):46–52, May 1996.

[801] Marilyn A. Walker, Jeanne Fromer, Giuseppe Di Fabbrizio, Craig Mestel, and Don Hindle. What can I say: Evaluating a spoken language interface to email. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 582–589, Los Angeles, CA, March 1998.

[802] Gregory K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34(4):31–44, April 1991.

[803] S. Wartick. Boolean operations. In W. B. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures & Algorithms*, pages 264–292. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[804] John A. Waterworth and Mark H. Chignell. A model of information exploration. *Hypermedia*, 3(1):35–58, 1991.

[805] WebTaxi: Main Page. http://www.webtaxi.com, 1998.

[806] S. Weibel, J. Godby, E. Miller, and R. Daniel. OCLC/NCSA Metadata Workshop Report: The Essential Elements of Network Object Description, March 1995. http://purl.oclc.org/oclc/rsch/metadataI.

[807] S. Weibel and E. Miller. Dublin Core Metadata, 1997. URL: http://purl.org/metadata/dublin_core.

[808] Andreas S. Weigend and Neil A. Gerschenfeld. *Time Series Prediction: Forecasting the Future and Understanding the Past*. Addison Wesley, 1994.

[809] Bella H. Weinberg. Bibliographic coupling: A review. *Information Storage and Retrieval*, 10(5-6):189–196, 1974.

[810] R. Weiss, B. Vélez, M. Sheldon, C. Nemprempre, P. Szilagyi, and D. K. Gifford. HyPursuit: A hierarchical network engine that exploits content-link hypertext clustering. In *Proc. of the 7th ACM Conference on Hypertext and Hypermedia*, pages 180–193, Washington, DC, USA, 1996.

[811] H. G. Wells. World Brain: The idea of a permanent world encyclopedia. Contribution to the New Encyclopedie Francaise, 1937. http://sherlock.berkeley.edu/wells/world_brain.html.

[812] H. D. White and K. W. McCain. Bibliometrics. *Annual Review of Information Science and Technology*, 24:119–186, 1989.

[813] Howard S. White, editor. *Library Technology Reports*, volume 33(5). American Library Association, Chicago, IL, September/October 1997.

[814] Robert Wilensky, Yigal Arens, and David N. Chin. Talking to UNIX in English: An overview of UC. *Communications of the ACM*, 27(6), 1984.

[815] R. Wilkinson and P. Hingston. Using the cosine measure in a neural network for document retrieval. In *Proc. of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 202–210, Chicago, USA, Oct 1991.

[816] Peter Willet. Recent trends in hierarchical document clustering: A critical review. *Information Processing and Management*, 24(5):577–597, 1988.

[817] J. Williams. *Bots and other Internet Beasts.* Prentice Hall, Englewood Cliffs, NJ, 1996.

[818] Michael David Williams. What makes rabbit run? *International Journal of Man-Machine Studies*, 21(4):333–352, 1984.

[819] W. Willinger and V. Paxson. Where mathematics meets the Internet. *Notices of the AMS*, 45(8):961–970, 1998.

[820] T. Winograd. Conceptual models for comparison of digital library systems and approaches. Stanford Digital Library Project Working Paper, July 1995. http://www-diglib.stanford.edu/diglib/WP/PUBLIC/DOC13.html.

[821] James A. Wise, James J. Thomas, Kelly Pennock, David Lantrip, Marc Pottier, and Anne Schur. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proc. of the Information Visualization Symposium 95*, pages 51–58. IEEE Computer Society Press, 1995.

[822] N. Wiseman. Implementing a national access management system for electronic services: Technology alone is not enough. *D-Lib Magazine*, March 1998. http://www.dlib.org/dlib/march98/wiseman/03wiseman.html.

[823] I. H. Witten, R. Neal, and J. G. Cleary. Arithmetic coding for data compression. *Communications of the ACM*, 30(6):520–541, 1987.

[824] Ian H. Witten, Craig Nevill-Manning, Roger McNab, and Sally Jo Cunningham. A public library based on full-text retrieval. *Communications of the ACM*, 41(4):71–75, 1998.

[825] I. H. Witten, A. Moffat, and T. C. Bell. *Managing Gigabytes: Compressing and Indexing Documents and Images.* Van Nostrand Reinhold, New York, 1994.

[826] Kent Wittenburg and Eric Sigman. Integration of browsing, searching, and filtering in an applet for Web information access. In *Proc. of the ACM Conference on Human Factors in Computing Systems, Late Breaking Track*, Atlanta, GA, USA, 1997. http://www1.acm.org:82/sigs/sigchi/chi97/proceedings/short-talk/kw.htm.

[827] D. Woelk, W. Kim, and W. Luther. An object-oriented approach to multimedia databases. In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, pages 312–325, Washington, DC, USA, June 1986.

[828] L. Wong. BioKleisli. http://corona.iss.nus.sg:8080/biokleisli.html.

[829] L. Wong. BioKleisli Architecture. http://sdmc.krdl.org.sg/kleisli/kleisli/Architecture.html.

[830] S. K. M. Wong and Y. Y. Yao. On modeling information retrieval with probabilistic inference. *ACM Transactions on Information Systems*, 13(1):39–68, 1995.

[831] S. K. M. Wong, W. Ziarko, V. V. Raghavan, and P. C. N. Wong. On modeling of information retrieval concepts in vector spaces. *ACM Transactions on Database Systems*, 12(2):299–321, 1987.

[832] S. K. M. Wong, W. Ziarko, and P. C. N. Wong. Generalized vector space model in information retrieval. In *Proc. 8th ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 18–25, New York, USA, 1985.

[833] A. Wood. DC-4: NLA/DSTC/OCLC Dublin Core Down Under / The 4th Dublin Core Metadata Workshop, March 1997. http://www.dstc.edu.au/DC4/.

[834] A. Woodruff, P. Aoki, E. Brewer, P. Gauthier, and L. Rowe. An investigation of documents from the World Wide Web. In *5th WWW Conference*, Paris, France, 1996.

[835] H. Wu and G. Salton. The estimation of term relevance weights using relevance feedback. *Journal of Documentation*, 37(4):194–214, 1981.

[836] S. Wu and U. Manber. Agrep — a fast approximate pattern-matching tool. In *Proc. of USENIX Technical Conference*, pages 153–162, San Francisco, CA, USA, 1992.

[837] S. Wu and U. Manber. Fast text searching allowing errors. *Communications of the ACM*, 35(10):83–91, October 1992.

[838] J. Xu and W. B. Croft. Query expansion using local and global document analysis. In *Proc. ACM-SIGIR Conference on Research and Development in Information Retrieval*, pages 4–11, Zurich, Switzerland, 1996.

[839] Yahoo!: Main page. http://www.yahoo.com, 1995.

[840] Byoung-Kee Yi, H. V. Jagadish, and Christos Faloutsos. Efficient retrieval of similar time sequences under time warping. In *ICDE'98*, pages 201–208, Orlando, Florida, Feb 1998.

[841] Degi Young and Ben Shneiderman. A graphical filter/flow model for Boolean queries: An implementation and experiment. *Journal of the American Society for Information Science*, 44(6):327–339, July 1993. With permission from University of Maryland, USA, Shneiderman & Young, 1993.

[842] C. T. Yu and G. Salton. Precision weighting — an effective automatic indexing method. *Journal of the ACM*, 23(1):76–88, January 1976.

[843] C. T. Yu, C. Buckley, K. Lam, and G. Salton. A generalized term dependence model in information retrieval. *Information Technology: Research and Development*, 2(4):129–154, October 1983.

[844] B. Yuwono and D. L. Lee. Search and ranking algorithms for locating resources on World Wide Web. In *Proc. of the Int. Conference on Data Engineering (ICDE)*, pages 164–171, New Orleans, USA, 1996.

[845] B. Yuwono and D. L. Lee. Server ranking for distributed text retrieval systems on the Internet. In *Proc. of the 5th Int. Conf. on Databases Systems for Advanced Applications (DASFAA)*, pages 41–50, Melbourne, Australia, 1997.

[846] L. A. Zadeh. Fuzzy sets. In D. Dubois, H. Prade, and R. R. Yager, editors, *Readings in Fuzzy Sets for Intelligent Systems*. Morgan Kaufmann, 1993.

[847] G. Zipf. *Human Behaviour and the Principle of Least Effort*. Addison-Wesley, 1949.

[848] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343, 1977.

[849] J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, 24(5):530–536, 1978.

[850] J. Zobel and A. Moffat. Adding compression to a full text retrieval system. *Software Practice and Experience*, 25(8):891–903, August 1995.

[851] Justin Zobel, Alistair Moffat, and Kotagiri Ramamohanarao. Inverted files versus signature files for text indexing. Technical Report CITRI/TR-95-5, Collaborative Information Technology Research Institute, Department of Computer Science, Royal Melbourne Institute of Technology, Australia, July 1995.

[852] Albert Y. Zomaya, editor. *Parallel and Distributed Computing Handbook*. McGraw-Hill, New York, 1996.

Note: some of the bibliography is only available on the Web, in particular conferences and electronic journals related to the Web itself or digital libraries. Most of these references are available on the Web page of the book (see Chapter 1).

Index

ad hoc retrieval 21, 437
adaptive compression models 177
addressing granularity 193
agent technologies 420
algebraic models 41
all-pairs or spatial-join query 437
all pairs queries 346
All4One 388
AltaVista 374, 384
Amdahl's law 437
AMIT system 300
animation 260
AOL Netfind 384
application programming interface (API) 152
architecture
  multiple instruction multiple data streams 230
  multiple instruction single data streams 230
  single instruction stream multiple data stream 230
  single instruction stream single data stream 230
  taxonomy of 230
arithmetic coding 175
ARJ 145
ASCII 144, 437
Ask Jeeves 288, 374
association clusters 125
AU 157
Audio 157
automatic feature extraction 360
automatic global analysis 131
automatic local analysis 123
automation 217
average precision at seen relevant documents 80
AVI 158

basic queries 100
Bayes' theorem 58
Bayesian belief networks 48
Bayesian networks 48, 60
  computational costs 60
  models 59, 61
belief networks 48, 437
  model 56
  probability space 56
  vector, ranking 59
Berkeley project 426
BibTeX 143
binary independence retrieval (BIR) 30
Binary Large OBjects (BLOBs) 329
binary weights 27
BioKleisli 423
bit-parallelism 218, 437
block addressing 193, 438
bookmark 267
Boolean models 20, 25, 438
  definition 26
  extended 38
  inference network 53
Boolean queries 102, 207, 279
Brown Corpus 147
browsing 65, 265
  flat 65

models 65
structure guided 66
browsing the Web 384, 386
directories 384
helpful tools 387
metasearchers 387
searching 386
brushing and linking 260
brute force algorithm 209

CA SEARCH 401
CACM 94
CACM collection 92, 438
canonical tree 181
Carnegie Mellon project 426
cascade style sheets (CSS) 152
Cat-a-Cone 319
cataloging 143
categories 269
hierarchies 268
categories for results set context 297
Cha-Cha intranet search system 299
Character Large OBject data 329
Cheshire II 411
child node 49
Chinese task in TREC-6 90
CISI collection 438
client/server architecture 331
clustering 123, 124, 438
algorithms 28
local 124
local analysis 123
techniques 268
co-citation clustering 276
coding 438
collections 267, 438
lists of 267
partitioning 251
color highlighting 260
command language 278
command lines 280
common command language 114
Compact Disk Read only Data exchange (CD-RDx) 114
complement 35
complete link algorithm 135
Compress 145
compression 222-4
inverted files 224
sequential searching 223
signature files 225
software 145
suffix arrays 225
suffix trees 225
of text 164, 438
compression ratio 176
Computer Graphics Metafile (CGM) 159
content-based queries 334, 439
content rating 143
ConText 330
context
categories for results set 297
document surrogates 289
heterogeneous Web site 299
Hyperlinks 299
InfoCrystal 294
KWIC 290
lattices 295
Lyberworld 295
Mappuccino 301
query term hits 289, 293
SeeSoft 292
SuperBook 296
tables 301
TileBars 291
VIBE 295
context queries 101
controlled vocabulary 170
conversion 439
Copernic 388
CORE project 297, 425
coverage ratio 83
crawlers 373
crawling 382
cross language task in TREC-6 90
CSTR project 424
Cystic Fibrosis Collection 94, 439

data blades 330

data cartridges 330, 439
data mining 439
data modelling 326, 328
  multimedia data support 329
  MULTOS Data Model 331
data quality on Web 368
data retrieval 1, 326, 439
database industry 439
database management systems (DBMS) 325, 329
  multimedia data support 329
database vendors 439
databases 399, 400
  iindexing 400
  for online retrieval 399
  producers 399, 439
  vendors 399
DejaNews 374
descriptive metadata 143
DIALOG 281, 398, 400, 404
dialogs in starting points 276
dictionary text compression 175, 183
Dienst reference implementation 424
digital libraries 7, 415, 440
  access 424
  architectural issues 418
  definitions 417
  distributed collections 422
  federated search 424
  interfaces 425
  international range of efforts 427
  metadata 430
  multilingual documents 420
  multimedia documents 421
  projects 425
  protocols and federation 429
  prototypes 425
  societies 416
  structured documents 421
  usability 428
  on Web 7
  Z39.50 429
Digital Libraries Initiative 425
Digital Libraries Initiative - Phase 2 (DLI-2) 427
Digital Library projects 412
digital objects 440
  in digital libraries 419
digital preservation 440
digital signatures 143
direct manipulation 278
directed acyclic graphs (DAGs) 48
DirectHit 374
director 269, 440
distance function 148, 353
  one-dimensional time series 353
distributed computing 440
  architectures 231
distributed data on web 368
distributed information retrieval 249, 252, 440
  collected partitioning 251
  query processing 253
  source selection 252
  World Wide Web 254
DLI 440
DLITE 319
DLITE interface 314
document collection 86
  TREC 86
document content 289
document partitioning 234, 237
  logical 236
  physical 237
  term 238
Document Style Semantic Specification Language 150
document surrogates 289, 441
document type declaration (DTD) 149
document type definition 152
documents 29, 78, 141, 142, 165, 173, 372, 420, 421, 440
  characteristics 142
  cutoff values 78
  index terms selection 169
  logical view 5
  multilingual 420
  multimedia 421

normalized frequency 29
preprocessing 165
query term hits 293
spoken, retrieval in TREC-6 90
stemming 168
structured 421
syntax 141
thesauri 170
on web 372
DTD 151, 441
DynaCat 297
dynamic programming 216
dynamic structure of hierarchical models 112

E measure 82, 441
EBCDIC 144
eBLAST 384
EchoSearch 388
edit distance 105, 441
efficiency 441
elastic windows 312
eLib 427
encryption 164
Endeavor Information Systems, Inc. 411
entropy 176, 441
in information theory 145
ENVISION project 425, 428
Envision system 301
EOD International (EOSi) 411
epistemological view of probability 50
ERCIM program 427
errors, allowing 105
exact match 441
Excite 374, 384
extended boolean model 441
extended patterns 106, 219, 441
searching for 105
eXtensible Markup Language 154
Extensible Style sheet Language (XSL) 156

faceted queries 281, 442
false drop 205
feature 442
feature extraction 353
in GEMINI 352
one-dimensional time series 353
federated search 442
file compression 183
inverted 184
filter-flow model 284
filtering 22, 442
retrieval 21
in TREC-6 90
fixed structure 108
flat browsing 65
FLI 158
focus-plus-context 260
form fillin 278
form-like fixed structure 107
formatted output specification instance 150
forms 280
free text queries 287
FREESTYLE 406
frequentist view of probability 50
frustration in evaluation 84
full text 442
document 24
fuzzy document sets 37
fuzzy information retrieval 35
fuzzy model 442
fuzzy set model 34
fuzzy set theory 35

generalized vector space model 442
assumption 41
GEneric Multimedia object INdexIng (GEMINI) 349, 360
algorithm 352
approach 348
automatic feature extraction 360
Glimpse 386
global analysis 442
Google 381
government information locator service 430
granularity 442

Graphic Interchange Format (GIF) 157
graphical evaluations 274
graphical overviews 274
graphics and virtual reality 159
grid-files 347
group relevance judgements 308
guided tour 443

hamming distance 148
handles in digital libraries 419
harmonic mean 82, 443
Harvest 375, 376
Heaps' law 147, 148, 193, 194, 371, 443
heterogeneous data on web 369
HiBrowse 271, 285
hierarchical structure 107, 109, 110
  sample 110
HotBot 374, 384
hotlist 267
HTML 152, 443
Huffman coding 175, 179, 180, 443
  byte-oriented 182
  compression 223
  methods 178
Huffman tree 181
human-computer interaction 258
  design principles 258
  interactive systems 261
human-computer interfaces (HCI) 443
H.W. Wilson 398, 400
hyperlinks 299, 392
  searching 392
hypertext 67
  model 66, 443
  structure 107, 108
Hypertext Induced Topic Search (HITS) 380, 381
HyperText Markup Language (HTML) 68, 152
Hypertext Transfer Protocol (HTTP) 68
HyTime 159

IBM digital library 427
icons 260
Illinois project 426
Illustra 330
index/indexing 191, 199, 205, 220, 443
  compressed 224
  construction 196
  databases 400
  definition 26
  generalized vector space model 42
  pattern matching 220
  signature files 205
  structure 199
  suffix arrays 199
  suffix trees 199
  text 199
  weighting *see* wighting
index points 199, 444
index terms 5, 163, 444
  see also keyword
  in automatic global analysis 133
  Boolean model 26
  definition 56
  dependencies 43
  independence of 443
  probabilistic model 31
  term relationships 172
  thesaurus 171
index weights, vector model 27
inference networks 48, 444
  assumption 51
  Boolean model 53
  combining evidential sources 55
  model 49
  prior probabilities 52
  tf-idf ranking strategies 54
InfoBus 420
InfoGrid layout 312
information access process 262
  browsing 265
  navigating 265
  non-search parts of 265
  querying 265
  scanning 265

information requests 88, 89
  relevant documents 89
information retrieval (IR) 444
  early developments 6
  model for 446
  performance 444
  practical issues 8
  process of 9
information theory 145
information visualization 260
information visualizer 319
informative feedback 444
  in human-computer interaction 258
Informedia 421
Infoseek 384
Innovative Interfaces, Inc. (III) 411
intellectual property rights 143
inter-cluster dissimilarity 29
interaction 262
interaction access process 266
interaction search formulation 127
interaction systems 261
  evaluating 261
  models of interaction 262
interactive tasks in TREC-6 90
interfaces 267, 304, 425
  context 289
  integrating scanning, selection, and querying 318
  relevance judgement 304
  search process 309
  starting points 267
  string matching 309
internal locus of control in HCI 259
interoperability 444
intersection of two 35
intra-cluster similarity 29
Intranet 444
inverse document frequency 29
inverted files 224, 445
  compression 224
  indices 383
  parallel IR 235
  SIMD architectures 245
ISI 94
  collection 93
  statistics 94

JBIG 157
Joint Photographic Experts Group (JPEG) 157

keyword-based querying 100
  Boolean 102
  context 101
  natural language 103
  single-word 100
keyword connection matrix 36
keyword-in-context (KWIC) 290
keywords 445
knowbots 373
Knuth-Morris-Pratt search algorithm 210
Kohonen's feature map 445
KWIC 445

latent semantic indexing 445
  model 44
lattices 295
Letizia 307
Levenshtein distance 105, 148, 445
lexical analysis of text 165
lexicographical order 445
LEXIS-NEXIS 267, 287, 317, 398, 400
libraries 7
library management systems 411
linear quadtrees 347
link structure 301
local analysis 445
local context analysis 129, 445
  query expansion 129
logical document partitioning 236
logical view of documents 445
LookSmart 384
loosely coupled MIMD systems 231
lower-bounding 353, 358
  one-dimensional time series 353
  two-dimensional color map 358
lower bounding lemma 351
Lyberworld 295

Lycos subjects 384

machine readable cataloging record 143, 409
Magellan 374, 384
magic lenses 260
Mandelbrot distribution 146
Mapuccino 301
MARC 430, 446
MARIAN 411, 425
markup languages 149
Mathematical Markup Language (MathML) 156
MEDLINE 143, 269, 401
MEDLINE database 94
MELVYL 411
Melvyl system 280, 319
memory load in HCI 259
menus 280
  selection 278
MeSHBrowse interface 270
metacrawler 387
metadata 141, 142, 143, 430, 446
  common forms 143
  digital libraries 430
Metafile for Interactive Documents (MID) 160
metasearch 446
  on Web 387
metric clusters 126
Microsoft 371, 374
MIDI 157
MIMD 446
  architectures 233
  parallel IR 233
MIMD parallel processor 249
MIME (Multipurpose Internet Mail Exchange) 144
MISD 446
modeling 19, 446
  definition 23
  natural language 145
  taxonomy 20
Moving Pictures Expert Group 158
multidimensional scaling (MDS) 446
multimedia 156
  data 447
  data conditions 335
  formats 157
  graphics and virtual reality 159
  hyTime 159
  operations 163
multimedia database management system 447
multimedia information retrieval system 325, 345, 447
  data modeling 328
  inverted files 330
  query languages 334
multitasking 233, 447
MULTOS data model 331
MULTOS query language 339
Murax system 288

n-gram 447
National Institute of Standards and Technology (NIST) 85
National Library of Medecine 398
natural language 103, 278, 287
  processing in TREC-6 90
navigating 265
navigational task in hypertext 67
nearest-neighbour queries 346, 447
Netscape 371
networked CS technical reference 424
Networked Digital Library of Theses and Dissertations (NDLTD) 428
neural networks 46, 447
New York Times 401
NewHoo 384
noise 163
non-binary weights 27
non-overlapping lists 62
  model 447
non-relevant document 27
normalized frequency 29
Northern Light 374
Object Database Management Group (ODMG) 328

object-oriented database management systems 328
object-relational database technology 447
occurrence list 447
occurrences 192, 195
  inverted file structure 192
  manipulation 195
  retrieval 195
OCLC 398, 400
OCR 448
Okapi 411
one-dimensional time series
  distance function 353
  experiments 355
  feature extraction 353
  lower-bounding 353
one-dimensional time series in multimedia IR 353
OneSeek 388
Online Computer Library Center (OCLC) 409
online public access catalog (OPACs) 408, 410, 448
  Alternatives to vendor 410
  content 408
  end users 410
  vendors and products 410
online public access catalogs 407
online retrieval systems 403, 404
  natural language searching 406
ontologies 143
OpenText 422
optical character recognition 448
optimization 326
Oracle DBMS 329
ordered inclusion 112
overlapped lists 110
overlapping windows 311, 317
  poor use of 317
overviews 268, 269, 272, 276
  automatically derived 272
  category 269
  co-citation clustering 276
  directory 269
PAD++ 426
PageRank 381
panning 448
panning and zooming 260
parallel computing 230, 448
  architectures 230
  performance measures 231
parallel information retrieval 232, 448
  inverted files 235
  MIMD architectures 233
  signature files 239
  suffix arrays 239
parallel multitasking 234
parents of a node 49
Patricia tree 199
pattern matching 104, 215, 220
  automation 217
  bit-parallelism 218
  dynamic programming 216
  filtering 218
  indexing 220
  inverted files 220
  regular expressions extended patterns 219
  string matching allowing errors 216
  suffix arrays 221
  suffix trees 221
  using indices 220
patterns 104, 448
performance evaluation 73
perspective wall 302
phrases 101
  and proximity search algorithm 215
portable document format (PDF) 144
portable network graphics (PNG) 157
Porter's algorithm 433
post-coordinate ranking 282
Postscript 141, 144
precision 75, 80, 81, 448
  appropriateness 81
  average 80

defined 75
histograms 80
in TREC-6 90
prefixes 105
prior probability distribution in inference networks 52
privacy levels 143
probabilistic models 20, 30, 120, 448
advantage 34
alternative 48
disadvantages 34
fundamental assumption 31
term reweighting 121
projects 425
Proteus 388
prototypes 425
proximal nodes 63, 111
model 449
proximity 101, 337
context 101
query languages 337
pseudo-document 45
PsycINFO 401

QBIC project 357
query answer 327
query by example 335
Query By Image Content (QBIC) 421
query expansion 118, 124, 129, 131, 134, 449
definitions 124
similarity thesaurus, based on 131
statistical thesaurus 134
through local context analysis 129
vector model 118
query iteration 327
query languages 99, 113, 334
conditions on multimedia data 335
multimedia IR 334
in MULTOS 339
proximity 337
request specification 335
uncertainty 337
weights in query expressions 337
query operations 117
term reweighting 120
user relevance feedback 118
query preview 285, 449
query processing 253
distributed IR 253
and optimization 326
query protocols 113
query specification 278, 282, 326
Boolean queries 279
from command lines to forms and menus 280
faceted queries 281
free text queries 287
graphical approaches 282
natural language 287
phrases and proximity 286
query syntax tree 102, 207
query term hits 289
between documents 293
querying 265, 318, 449
QuickTime 158
quorum-level ranking 282

R-Precision 80
$R*$-trees 347, 355
range query 449
ranges 105
ranking belief network vector 59
recall 75, 81, 449
appropriateness 81
defined 75
effort 84
and precision 75
reference collections 84, 449
references, lists of 111
regular expressions 105, 219, 449
searching for 105
relative recall 84
relevance feedback 65, 118
relevance judgements 303, 304, 305, 308
fetching relevant information 307

group 308
interfaces 304
pseudo-relevance feedback 308
user interaction 305
relevant document 27
repositories 449
in digital libraries 419
request specification 335
resource description format 156
resource description framework (RDF) 143, 431
retrieval evaluation 73
retrieval performance evaluation 73, 74
retrieval task 450
retrieval unit 100, 450
Rich Text Format (RTF) 144
robots 373
rooms 319
routing 22

sample space 57
belief network model 58
Santa Barbara project 426
SavvySearch 387
scalar clusters 127
scanning 265, 318
scatter/gather 450
scenarios in digital libraries 416
search algorithm 351
search broker 374
search engines 380
crawling the Web 382
distributed architecture 375
indices 383
ranking 380
user interfaces 377
on the Web 373
search history 450
search process 309
search history 317
string matching 309
window management 311
Search Spaniel 388
search starting point 450
searching 195, 202, 206, 220, 386, 392
algorithm 210-15
Boyer-Moore family algorithm 211
browsing the Web 386
brute force algorithm 209
compression 222
dynamic 393
hyperlinks 392
inverted index 195
Knuth-Morris-Pratt 210
natural language 406
phrases and proximity 215
practical comparison 214
practical comparison search algorithm 214
query languages 392
sequential 209
Shift-Or 212
Shift-Or search algorithm 212
signature files 206
structural queries 222
suffix automation 213
suffix automation search algorithm 213
suffix tree 202
SeeSoft 292
selection 318
semi-static compression models 177
semi-static text 450
semi-structured data 450
SenseMaker system 302, 428
sequencing 66
sequential inverted files 220
sequential or on-line text searching 450
sequential searching 209
compression 223
set theoretic alternative 34
set theoretic models 20
signature files 205, 206, 207, 225, 239, 240, 451
compression 225
construction 207
indexing 205

parallel IR 239
searching 206
SIMD architectures 240
structure 205
similarity models 148
similarity thesaurus 131
definition 131
single instruction multiple data (SIMD) 451
architectures 240, 245
inverted files 245
signature files 240
single value summaries 79
single-word queries 100
SISD 451
SketchTrieve interface 316
Smart system 118
SMP (Symmetric Multiprocessor) 451
software agents for searching 393
SONIA system 298
source selection 451
spaceless words 178, 180
spaces in digital libraries 416
spatial access methods 347, 451
spatial joins 346
speedup 451
spiders 373
spoken documents, retrieval in TREC-6 90
spread activation 46
SQL-92 330
SQL3 Query Language 338, 451
Standard Generalized Markup Language (SGML) 142, 149, 450
tags 150
Standard Music Description Language (SMDL) 160
Stanford proposal for internet meta-searching (STARTS) 251, 451
starting points 267
automated source selection 278
dialogs and wizards in 276
lists of collections 267
overviews 268
statistical text compression 175, 452
statistical thesaurus 134
stemming 5, 168, 452
stopwords 100, 146, 167, 452
elimination of 167
STR models
non-overlapping lists 62
proximal nodes 63
streams in digital libraries 416
string matching allowing errors 216
structural queries 106, 222
structure 112
restrictions 112
static 112
structure-guided browsing 66
Structured Full-text Query Language (SFQL) 114
structured text retrieval models 61
structures in digital libraries 416
sub-pattern match 346
substrings 105
suffix arrays 199, 200, 201, 203, 221, 225, 239
compression 225
construction 203
large texts 203
parallel IR 239
suffix trees 199, 200, 202, 203, 221, 225
compression 225
construction in main memory 203
searching 202
and suffix array 452
suffixes 105
summary table statistics 80
SuperBook 296, 299, 311
Web 299
SuperBook layout 313
Sybase SQL server 330
Synchronized Multimedia Integration Language (SMIL) 156
syntax tree 452
Syskill and Webert 307

table lens 302
tag 452
Tagged Image File Format (TIFF) 157
TARGET 406
TCP/IP 250
term partitioning 238
term reweighting 118, 120
  vector model 118
term-term correlation matrix 36
term-to-term independence 452
text 112, 141, 144, 163
  formats 144
  indexing 199
  information theory 145
  integration 112
  languages and protocols 141
  lexical analysis 165
  modeling natural language 145
  similarity models 148
text compression 173, 186
  basic concepts 175
  coding 179
  comparing 186
  inverted file structure 192
  inverted files 192
  modeling 177
  motivation 173
  statistical methods 176
  techniques 186
text encoding initiative (TEI) 151, 431
text operations 5
Text REtrieval Conference (TREC) 84, 85
  Benchmark 89
  document collection 86, 453
  evaluation measures 91
text structure 452
textual image 452
thesaurus 170, 453
  clustering 173
  index terms 171
  in IR 172
  use of 172
tightly coupled MIMD systems 231
TileBars 291
tiled windows 311
TIPSTER program 85
tree matching 112
Truevision Targa image file (TGA) 157
two-dimensional color images 357
  distance functions 357
  image features 357
two-dimensional color maps 358
  experiments 360
  lower-bounding 358

uncertainty 337
  query languages 337
unicode 144, 421
Uniform Resource Identifier (URI) 143
Uniform Resource Locator (URL) 143, 453
union of two 35
unordered inclusion 112
USENET 374
user, tasks 4
user information need 1, 453
user interaction 305
user interfaces 377
  search engine 377
user-oriented measures of evaluation 83
user profile 22
user relevance feedback 118, 453
  feedback strategies 122
USMARC 430

vector model 118, 453
  advantages 30
  definition 27
  query expansion 118
  term reweighting 118
vector models 20
Venn diagrams 283
very large corpus task in TREC-6 90
VIBE 295

Virtual Reality Modeling Language (VRML) 159
VISAGE 318
visualization, role of 259
vocabulary 147, 174, 192, 453
  search 195
volatile data on web 368
VQuery system 283

WAIS 453
walkers 373
wanderers 373
WATERS initiative 424
WAVE 157
Web Query Languages 392
WebCompass 388
WebCrawler Select 384
WebCutter system 302
WebFerret 388
WebGlimpse 386
WebLog 392
WebQuery 380
WebSeeker 388
WebSQL 392
WebTOC 300
weighting
  definition 25
  generalized vector space model 42
  in query expressions 337
  query language 337
  vector model 27, 28
WESTLAW 406
whole match 346
wide area information service 113
WIN 406
Window management 311
wizards in starting points 276
word 101
word-based compression models 177
words 105, 147
  average length 147
working memory load 454
World Wide Web 1, 68, 143, 254, 367
  browsing 384
  centralized architecture 373
  characterizing 369
  crawling the 382
  distributed IR 254
  languages used 371
  measuring 369
  modeling 371
  search engines 373
  search examples 390
  searching 367
  taxonomy of languages 161
  teaching user 391
  user problems 389
WQL 392
W3QL 392
WSQ 157

XML 454

Yahoo 269, 371, 384

Z39.50 423, 454
ZIP 145
Zipf's Law 146, 371, 454
zooming 454