

Information Retrieval im WWW – Literaturliste (Sommer 2001)

1. Die Architektur der Suchmaschine Google

- S. Brin, L. Page: The Anatomy of a Large-Scale Hypertextual Search Engine. Proceedings of the 7th International World Wide Web Conference (WWW7), 1998.

2. Aufbau eines Web-Index zur effizienten Suche

- R. Baeza-Yates, B. Ribeiro-Neto:
Modern Information Retrieval. ACM Press, 1999. Kapitel 4 (Query Languages), Kapitel 6.3 (Text), Kapitel 7.2 (Document Preprocessing), und Kapitel 8 (Indexing and Searching).

3. Die Graphenstruktur und Evolution des WWW

- A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener:
Graph structure in the web. Proceedings of the 9th International World Wide Web Conference (WWW9), 2000.
- J. Cho, H. Garcia-Molina:
The Evolution of the Web and Implications for an incremental Crawler. Proceedings of 26th International Conference on Very Large Databases (VLDB), September 2000.

4. PageRank: Anfrageunabhängige Bewertung von Webseiten

- L. Page, S. Brin, R. Motwani, T. Winograd:
The PageRank Citation Ranking: Bringing Order to the Web. Manuscript.
- T. Haveliwala:
Efficient Computation of PageRank. Technical Report, Stanford University, 1999.

5. HITS: Anfrageabhängige Bewertung von Webseiten

- J. Kleinberg:
Authoritative Sources in a Hyperlinked Environment. Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms, 1998. Extended version in Journal of the ACM 46(1999).
- D. Gibson, J. Kleinberg, P. Raghavan:
Inferring Web Communities from Link Topology. Proceedings of the 9th ACM Conference on Hypertext and Hypermedia, 1998.

6. Webcrawler

- J. Cho, H. Garcia-Molina, L. Page:
Efficient Crawling Through URL Ordering. Proceedings of the 7th World Wide Web conference (WWW7), 1998.
- J. Cho, H. Garcia-Molina:
The Evolution of the Web and Implications for an incremental Crawler. In Proceedings of 26th International Conference on Very Large Databases (VLDB), September 2000.

7. Einführung in Information Retrieval-Modelle

- R. Baeza-Yates, B. Ribeiro-Neto:
Modern Information Retrieval. ACM Press, 1999. Kapitel 2 (Modeling).
- P. Raghavan:
Information Retrieval Algorithms: A Survey. In: Proceedings of SODA 1997, pp. 11–18.

8. Klassifikation von Webseiten

- S. Chakrabarti, B. Dom, P. Indyk:
Enhanced hypertext categorization using hyperlinks. Proceedings of ACM SIGMOD international conference on Management of data, 1998, pp. 307–318.
- Soumen Chakrabarti, Byron Dom, Rakesh Agrawal, Prabhakar Raghavan:
Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies. VLDB Journal, August 1998.

9. Erkennung von fast identischen Webseiten und Hosts

- A. Broder, S. Glassman, M. Manasse, G. Zweig:
Syntactic Clustering of the Web. Proceedings of the 6th International World Wide Web Conference (WWW6), 1997.
- A. Broder:
On the resemblance and containment of documents. In: Compression and Complexity of Sequences (SEQUENCES'97), pp. 21-29. IEEE Computer Society, 1998.
- J. Cho, N. Shivakumar, H. Garcia-Molina:
Finding replicated Web collections. In Proceedings of 2000 ACM International Conference on Management of Data (SIGMOD), May 2000.

10. Das Auffinden neuer Cyber-Communities

- R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins:
Trawling the web for emerging cyber-communities. Proceedings of the 8th International World Wide Web Conference (WWW8), 1999.
- R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins:
Extracting large-scale knowledge bases from the web. Proceedings of 25th International Conference on Very Large Data Bases (VLDB'99), 1999.

11. Gruppierung von Query-Resultaten

- O. Zamir, O. Etzioni:
Web document clustering: a feasibility demonstration. Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, 1998, pp. 46-54.

12. Vergleich und Bewertung von Suchmaschinen

- K. Bharat, A. Broder:
A technique for measuring the relative size and overlap of public Web search engines. Proceedings of the 7th International World Wide Web Conference (WWW7), 1998.
- M. Henzinger, A. Heydon, M. Mitzenmacher, M. Najork:
Measuring Index Quality Using Random Walks on the Web. Proceedings of the 8th International World Wide Web Conference (WWW8), 1999, pp. 213-225.

13. Studien über das Benutzerverhalten

- B. Jansen, A. Spin, J. Bateman, T. Saracevic:
Real life information retrieval: a study of user queries on the Web. SIGIR Forum, Vol. 32, No. 1, 1998, pp. 5–17.
- C. Silverstein, M. Henzinger, J. Marais, M. Moricz:
Analysis of a very large AltaVista query log. SRC-Technical Note 1998-014.
- C.W. Choo, B. Detlor, D. Turnbull:
Information Seeking on the Web: An Integrated Model of Browsing and Searching. 1999 ASIS Annual Meeting Contributed Paper.