Hauptseminar
Inside Google - Algorithmik von Suchmaschinen

Prof. Dr. E.W. Mayr

Wintersemester 2007

Themenliste

1. Power-Law Graphen
   Literatur: [Mit04]
2. Caching Strategien
   Literatur: [KNPW02]
3. Datenaggregation
   Literatur: [BM02]
4. Graphkomprimierung
   Literatur: [AM01]
5. Indexstrukturen für exakte Volltextsuche
   Literatur: [Goo02], Kapitel 9.2
6. Indexstrukturen für fehlertolerante Volltextsuche
   Literatur: [NBYST01]
7. Suchen im Hypertext
   Literatur: [ALL97, Nav98]
8. Clustering von Dokumenten
   Literatur: [ZE98]
9. Page Rank
   Literatur: [BP98]
10. Web Communities
    Literatur: [KRRT99]

Literatur

[ALL97] A. Amir, M. Lewenstein, and N. Lewenstein. Pattern matching in hypertext. In F. K. H. A. Dehne, A. Rau-Chaplin, J.-R. Sack, and R. Tamassia, editors, *Proceedings of the5th Workshop on Algorithms and Data Structures*, number 1272, pages 160–173, Halifax, Nova Scotia, Canada, 1997. Springer-Verlag, Berlin.

[AM01] Micah Adler and Michael Mitzenmacher. Towards compressing web graphs. In *DCC '01: Proceedings of the Data Compression Conference (DCC '01)*, page 203, Washington, DC, USA, 2001. IEEE Computer Society.

[BM02] Andrei Broder and Michael Mitzenmacher. Optimal plans for aggregation. In *PODC '02: Proceedings of the twenty-first annual symposium on Principles of distributed computing*, pages 144–152, New York, NY, USA, 2002. ACM Press.

[BP98] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30(1-7):107–117, 1998.

[Goo02] *Algorithm Design.* John Wiley & Sons, Inc., 2002.

[KNPW02] Bala Kalyanasundaram, John Noga, Kirk Pruhs, and Gerhard J. Woeginger. Caching for web searching. *Algorithmica*, 33(3):353–370, 2002.

[KRRT99] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Trawling the Web for emerging cyber-communities. *Computer Networks (Amsterdam, Netherlands: 1999)*, 31(11–16):1481–1493, 1999.

[Mit04] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1:226–251, 2004.

[Nav98] Gonzalo Navarro. Improved approximate pattern matching on hypertext. *Lecture Notes in Computer Science*, 1380:352–??, 1998.

[NBYST01] G. Navarro, R. Baeza-Yates, E. Sutinen, and J. Tarhio. Indexing methods for approximate string matching. *IEEE Data Engineering Bulletin*, 24(4):19–27, 2001. Special issue on Text and Databases. Invited paper.

[ZE98] Oren Zamir and Oren Etzioni. Web document clustering: A feasibility demonstration. In *Research and Development in Information Retrieval*, pages 46–54, 1998.