# Focused Crawling Bibliography

Ioannis Partalas, Georgios Paliouras, Ioannis Vlahavas
{partalas,vlahavas}@csd.auth.gr
paliourg@iit.demokritos.gr

May 6, 2008

## Focused Crawling Bibliography

[De Bra and Post, 1994], [Hersovici et al., 1998], [Cho et al., 1998], [Diligenti et al., 2000], [Grigoriadis and Paliouras, 2004], [Rennie and McCallum, 1999], [C.-C. Hsu, 2006], [Chakrabarti et al., 1999], [Page et al., 1998], [O'Meara and Patel, 2001], [Johnson et al., 2003], [Pant and Menczer, 2002], [Chakrabarti et al., 2002], [Liu et al., 2004],[Srinivasan et al., 2005], [Pant and Srinivasan, 2005], [Pant and Srinivasan, 2006], [Almpanidis et al., 2007], [Babaria et al., 2007], [Partalas et al., 2008].

## References

G. Almpanidis, C. Kotropoulos, and I. Pitas. Combining text and link analysis for focused crawling-an application for vertical search engines. *Information Systems*, 32(6):886–908, 2007.

Rashmin Babaria, J. Saketha Nath, Krishnan S, Sivaramakrishnan K R, Chiranjib Bhattacharyya, and M. N. Murty. Focused crawling with scalable ordinal regression solvers. In *24th international conference on Machine learning*, pages 57–64, 2007. ISBN 978-1-59593-793-3.

F. Wu C.-C. Hsu. Topic-specific crawling on the web with the measurements of the relevancy context graph. *Information Systems*, 31(4-5):232–246, 2006.

Soumen Chakrabarti, Martin van den Berg, and Byron Dom. Focused crawling: a new approach to topic-specific Web resource discovery. *Computer Networks*, 31(11–16):1623–1640, 1999.

Soumen Chakrabarti, Kunal Punera, and Mallela Subramanyam. Accelerated focused crawling through online relevance feedback. In *WWW '02: Proceedings of the 11th international conference on World Wide Web*, pages 148–159, 2002. ISBN 1-58113-449-5.

Junghoo Cho, Hector García-Molina, and Lawrence Page. Efficient crawling through URL ordering. *Computer Networks and ISDN Systems*, 30(1–7): 161–172, 1998.

P. M. E. De Bra and R. D. J. Post. Information retrieval in the World-Wide Web: Making client-based searching feasible. *Computer Networks and ISDN Systems*, 27(2):183–192, 1994.

Michelangelo Diligenti, Frans Coetzee, Steve Lawrence, C. Lee Giles, and Marco Gori. Focused crawling using context graphs. In *26th International Conference on Very Large Databases, VLDB 2000*, pages 527–534, 2000.

A. Grigoriadis and G. Paliouras. Focused crawling using temporal difference-learning. In *Proc. 3th Hellenic Conference on Artificial Intelligence*, pages 142–153, 2004.

Michael Hersovici, Michal Jacovi, Yoelle S. Maarek, Dan Pelleg, Menanchem Shtalhaim, and Sigalit Ur. The shark-search algorithm. an application: tailored web site mapping. In *Proceedings of the seventh international conference on World Wide Web*, pages 317–326, 1998.

J. Johnson, K. Tsioutsiouliklis, and C. L. Giles. Evolving strategies for focused web crawling. In *20th International Conference on Machine Learning (ICML-2003)*, 2003.

Hongyu Liu, Evangelos Milios, and Jeannette Janssen. Probabilistic models for focused web crawling. In *6th annual ACM international workshop on Web information and data management*, pages 16–22, 2004. ISBN 1-58113-978-0.

Tadhg O'Meara and Ahmed Patel. A topic-specific web robot model based on restless bandits. *IEEE Internet Computing*, 5(2):27–35, 2001.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

G. Pant and F. Menczer. MySpiders: Evolve your own intelligent Web crawlers. *Autonomous Agents and Multi-Agent Systems*, 5(2):221–229, 2002.

Gautam Pant and Padmini Srinivasan. Learning to crawl: Comparing classification schemes. *ACM Transactions on Information Systems*, 23(4):430–462, 2005.

Gautam Pant and Padmini Srinivasan. Link contexts in classifier-guided topical crawlers. *IEEE Transactions on Knowledge and Data Engineering*, 18(1):107–122, 2006. ISSN 1041-4347.

Ioannis Partalas, Georgios Paliouras, and Ioannis Vlahavas. Reinforcement learning with classifier selection for focused crawling. In *19th European Conference on Artificial Intelligence (accepted for presentation)*, pages X–X, 2008.

J. Rennie and A. McCallum. Efficient web spidering with reinforcement learning. In *16th International Conference on Machine Learning'*, pages 335–343, 1999.

P. Srinivasan, F. Menczer, and G. Pant. A general evaluation framework for topical crawlers. *Information Retrieval*, 8(3):417–447, 2005.