AcademicArticles

Nutch code and tools have been used in several academic projects. This page exists to track published papers that have used Nutch as well as workshops and conferences that might be of interest for nutch developers.

Known papers in Computer Science and other academic fields: oldest to newest

1. **A Search Engine for Natural Language Applications.** Michael J. Cafarella, Oren Etzioni. Proceedings of the 14th International World Wide Web Conference (WWW 2005).

2. **Thwarting the nigritude ultramarine: learning to identify link spam.** Isabel Drost and Tobias Scheffer. Proceedings of the 16th European Conference on Machine Learning (ECML 2005).

3. **Query Chains: Learning to Rank from Implicit Feedback.** Filip Radlinski and Thorsten Joachims. Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM, 2005.

4. **Focused Crawling Implementation** Rashmin, Saketh, Dr. Chiranjib , Shivaramakrinsna and Krishi. Large Scale Data Mining Research Group @ Machine Learning Lab, The Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India

5. **A Knowledge-directed Information Retrieval and Management Framework for Energy Performance Building Regulations** Lewis John McGibbney and Bimal Kumar. Proceedings of the ASCE International Workshop on Computing in Civil Engineering: June 19-22, 2011, Miami, FL, USA

6. **The Wombra Project: A Web-based Ontology-enhanced Multi-purpose Building Regulation Retrieval Application for Scottish Technical Standards** Lewis John McGibbney and Bimal Kumar. Proceedings of the CIB W78-W102 2011: International Conference –Sophia Antipolis, France, 26-28 October

Related workshops and conferences:

Adversarial Information Retrieval on the Web ... A workshop that took place at the 14th International World Wide Web Conference. It covers papers that present "the latest results in adversarial web IR, and address topics such as web spam, blog spam, cloaking, redirection, link optimization for PageRank, automated link spam detection, link bombs, reverse engineering of ranking algorithms, and propaganda."