

Web Intelligence 2005

Compiègne, France

held jointly with the Intelligent Agent Technology (IAT) conference

Martin Labský

*Dept. of Knowledge Engineering,
University of Economics Prague*

Topics

- World Wide Wisdom Web (W4)
- Social Networks and Social Intelligence
- Knowledge Grids and Grid Intelligence
- Web Mining and Farming
- Semantics and Ontology Engineering
- Web Agents
- Web Services
- Web Information Filtering and Retrieval
- Intelligent Human-Web Interaction
- Web Support Systems
- Intelligent E-Technology

Some frequent topics

- **web usage mining and page recommendation systems**
- **(un)supervised document clustering**
- computing web page popularity – page rank modifications
- web page annotation and wrappers

“IE from web” papers

- Srinivas Vadrevu, Saravanakumar Nagarajan, Fatih Gelgi, and Hasan Davulcu: **Automated Metadata and Instance Extraction from News Web Sites** [38]
- Pu-Jen Cheng, Hsin-Chen Chiao, Yi-Cheng Pan, and Lee-Feng Chien: **Annotating Text Segments in Documents for Search** [317]
- Martin Labský, Vojtech Svátek, Ondrej Šváb, Pavel Praks, Michal Krátký, and Václav Snášel: **Information Extraction from HTML Product Catalogues: from Source Code and Images to RDF** [401]

Interesting papers #1

- Jing Bai, Jian-Yun Nie, and G. Cao: **Integrating Compound Terms in Bayesian Text Classification** [598]
 - using both unigrams and phrases as naive bayes features
 - cond. probabilities for phrases smoothed with unigram probs
- Marco Maggini, Leonardo Rigutini, and Bing Liu: **An EM based training algorithm for Cross-Language Text Categorization** [529]
 - 2 approaches to multi-lingual document classification
- Seung Yeol Yoo and Achim Hoffmann: **Pseudo-Relevance Feedback in Web information Retrieval Using Segments' Subjective Importance Values** [422]
 - query adaptation by word sense disambiguation using Wordnet's context terms

Interesting papers #2

- Chih How Bong, Ting Kiong Wong, and N. Kulathuramaiyer: **Categorical Term Descriptor: A Proposed Term Weighting Scheme for Feature Selection** [313]
 - discussion of feature selection methods for document classification
- Eytan Adar and Lada Adamic: **Tracking Information Epidemics in Blogspace** [207]
 - tracking topic spreading across a large number of blog websites
- Juan Raposo, Alberto Pan, Manuel Álvarez, and Justo Hidalgo: **Automatically Generating Labeled Examples for Web Wrapper Maintenance** [250]
- Tie-Yan Liu and Wei-Ying Ma: **Webpage Importance Analysis using Conditional Markov Random Walk** [515]
- Paolo Massa and Conor Hayes: **Page-reRank: using Trusted Links to Re-rank Authority** [614]