

## Selected Topics in Web Information Retrieval and Mining

Advanced Course  
Winter Semester 2003/2004

Gerhard Weikum  
weikum@mpi-sb.mpg.de

[http://www.mpi-sb.mpg.de/units/ag5/teaching/ws03\\_04/sir-ws0304.htm](http://www.mpi-sb.mpg.de/units/ag5/teaching/ws03_04/sir-ws0304.htm)  
[http://www-dbs.cs.uni-sb.de/lehre/ws03\\_04/sir-ws0304.htm](http://www-dbs.cs.uni-sb.de/lehre/ws03_04/sir-ws0304.htm)

Winter Semester 2003/04

Selected Topics in Web IR and Mining

0-1

## Course Organization

- **Lectures:** Thursday 11-13 in 45/002 (beginning Oct 30)  
Office hours Prof. Weikum: Tuesday 14-15 in 46/401 (or appointment by e-mail)
- **Assignments & Tutoring Group:**  
Friday 14-16 in 46/024 (beginning Nov 7)  
Teaching assistants:  
Hanglin Pan (pan@mpi-sb.mpg.de)  
Sergej Sizov (sizov@mpi-sb.mpg.de)  
Stefan Siersdorfer (siersdorfer@mpi-sb.mpg.de)
- **Requirements for obtaining 6 credit points:**
  - passing an oral exam at the end of the semester
  - at least one demonstration of solution to paper assignment in group meeting,
  - each additional presentation of a good solution earns a bonus point, up to 3 points
  - bonus points improve the final grade by 0.33 per point
- **Note:**
  - prior knowledge in information retrieval is needed (see the course „Information Retrieval and Data Mining“, <http://www-dbs.cs.uni-sb.de/~irdm02/> or the book Soumen Chakrabarti: „Mining the Web: Analysis of Hypertext and Semistructured Data“, Morgan Kaufmann, 2002

Winter Semester 2003/04

Selected Topics in Web IR and Mining

0-2

## Tentative Outline

### Part 1: Personalized Information Search

- 1) Thu, Oct 30 - Topic-specific Authority Ranking
- 2) Thu, Nov 6 - Personalized Authority Ranking
- 3) Thu, Nov 13 - Query-Log-based Personalization
- 4) Thu, Nov 20 - Relevance Feedback

### Part 2: Efficient Ranking

- 5) Thu, Nov 27 - Index Pruning
- 6) Thu, Dec 4 - Rank Aggregation
- 7) Thu, Dec 11 - Top k Queries on Structured Data and Deep Web Sources

### Part 3: System Architecture

- 8) Thu, Dec 18 - Topic Distillation
- 9) Thu, Jan 8 - XML Search Engines
- 10) Thu, Jan 15 - Peer-to-Peer Search

### Part 4: Statistical Language Models and Topic Spaces

- 11) Thu, Jan 22 - Latent Semantic Spaces
- 12) Thu, Jan 29 - Unified Link and Content Models
- 13) Thu, Feb 5 - Topic Mappings

Winter Semester 2003/04

Selected Topics in Web IR and Mining

0-3