

2V+1Ü Information Retrieval and Text Mining

Sommersemester 2009

Hinrich Schütze

Schedule

- Mondays, 15:45–17:15, V38.03 (a total of 6 Mondays, see course home page)
- Tuesdays, 15:45–17:15, V38.03 (every Tuesday)

Instructor

Hinrich Schütze ([schuetze at ims.uni-stuttgart.de](mailto:schuetze@ims.uni-stuttgart.de)),
Office hours: Fridays, 13:15–14:15, office 1.14, IMS, Azenbergstr. 12

Credit

- To get credit for the course you have to turn in the assignments and get at least 50% of available points.
- To get credit for the course you have to solve the Moodle exercises for each lecture. This should be quick and you have an unlimited number of attempts.
- There will be an oral exam at the end of the semester. Topic: the assignments.
- If the number of students in the class exceeds 40, the oral exam will be replaced with a written exam.

Moodle

Please sign up for an account at <http://ir.ims.uni-stuttgart.de>. **Please use an account name that starts with the prefix “iir-” and contains your first name and your last name.** Example: “iir-hinrichschuetze”

Course home page

<http://www.ims.uni-stuttgart.de/Lehre/teaching/2009-SS/ir/>

Or go to Hinrich Schütze’s home page at

<http://www.ims.uni-stuttgart.de/~schuetze/>

and click on “Teaching”.

Course overview

- Boolean retrieval
- The term vocabulary and postings lists
- Dictionaries and tolerant retrieval
- Index construction
- Index compression
- Scoring, term weighting and the vector space model
- Computing scores in a complete search system
- Evaluation in information retrieval
- Text classification and Naive Bayes
- Vector space classification
- Flat clustering
- Hierarchical clustering
- Web search basics
- Web crawling and indexes
- Link analysis

Textbook

- Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze: *Introduction to Information Retrieval*, Cambridge University Press, 2008.

Hinweis für Studierende der Computerlinguistik

Der Schein kann entweder für das Hauptfach Linguistik (Diplom) oder für das Nebenfach Informatik angerechnet werden.