Information Retrieval Exercises

Introduction

Mario Sänger (saengema@informatik.hu-berlin.de)

General structure

- ## We will build groups of two students each
  - Each group has to solve 4/5 exercises
  - All exercises must be solved by all groups!
  - To solve each exercise, you'll get 2-3 weeks
  - Weeks in between are optional Q/A sessions

- ## "Solving" means: a Java implementation or some configuration / extension of Lucene
  - Assignments are implementation-heavy
  - No paper work!

General structure

- (Partial) Solutions are presented by groups in the form of 5-10 minutes talks
  - Each group has to present at least once during the course
  - You'll be able to pick when and what you'd like to present
  - If you don't decide yourself, you'll be assigned a presentation slot

- But: You should be able to explain your solution anyway

Tentative schedule

- ## Today: group formation; assignment 1
  - Crawl IMDB and implement a set of queries

- ## 14/16 May: presentation of solutions; assignment 2
  - Implement fast Boolean search in a 100 MB corpus

- ## 04/06 June: presentation of solutions, assignment 3
  - Use Lucene for Boolean information retrieval

- ## 25/27 June: presentation of solutions, assignment 4
  - Implement synonym expansion with Lucene

- ## 16/18 July: presentation of solutions, final ceremony

Competition (optional)

- Most assignments can be solved more or the less well
  - "Good" usually means: very fast

- The best groups in each assignment will get points
  - Best: 5 points, second: 3 points; third: 1 point
  - Extra points for creative solutions possible!

- The overall best group will get a little present at the end of the semester

Group formation

- Form groups of two students until the end of this week
  - You can choose a fancy group name ☺
  - Write me an email with the group partners + group name

- I provide a Google Sheet for group formation
  - Link: https://bit.ly/2HOaEpa

- I don't accept submissions by lonely individuals!

Additional information

- Information, slides and materials of the exercise
https://hu.berlin/ire18

- If you have questions about the assignments - contact me!
  – Come by or write me an email before you do something wrong!

- Email: saengema@informatik.hu-berlin.de
  – Office: Rudower Chaussee 26, IV.404

Questions?