Information Retrieval Exercises

Introduction

Samuele Garda (gardasam@informatik.hu-berlin.de)
General structure

• We will build groups of three students each
  – Each group has to solve 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 QA sessions

• “Solving” means:
  – Java implementation
  – No paper work!
# Tentative schedule

<table>
<thead>
<tr>
<th>DATE</th>
<th>CONTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Today</td>
<td>• Introduction</td>
</tr>
<tr>
<td></td>
<td>• Assignment 1: IMDB Spider and query scraped data</td>
</tr>
<tr>
<td>11/12 May</td>
<td>• Presentation solution Assignment 1</td>
</tr>
<tr>
<td></td>
<td>• Assignment 2: Boolean search for large corpus</td>
</tr>
<tr>
<td>01/02 June</td>
<td>• Presentation solution Assignment 2</td>
</tr>
<tr>
<td></td>
<td>• Assignment 3: Boolean IR with Lucene</td>
</tr>
<tr>
<td>15/16 June</td>
<td>• Assignment 4: Query expansion with Lucene</td>
</tr>
<tr>
<td>22/23 June</td>
<td>• Presentation solution Assignment 3</td>
</tr>
<tr>
<td>29/30 June</td>
<td>• Assignment 5: Co-locations</td>
</tr>
<tr>
<td>06/07 July</td>
<td>• Present solution Assignment 4</td>
</tr>
<tr>
<td>15/16 July</td>
<td>• Presentation solution Assignment 5</td>
</tr>
</tbody>
</table>
General structure

• 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 always be able to explain your solution
Competition (optional)

• To make things more fun there will be a competition across groups for the best implementation

• The competition is optional and it has nothing to do with your final grade

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

• “Best” usually means very fast
Group formation

• Form groups **by the end of this week:**
  - You can choose a fancy group name
  - Write me an email with the group partners + group name

• I don’t accept submissions by lonely individuals

• See Moodle for helpers in group creation
Materials

• Information, slides and materials of the exercise will be on **Moodle**:
  - https://moodle.hu-berlin.de/course/view.php?id=104764

• I will update materials as well on the WBI group website:
  - https://www.informatik.hu-berlin.de/de/forschung/gebiete/wbi/teaching/archive/SS21/ue_ir/
Contact

• If you have any question do not hesitate to contact me!

• Email: gardasam@informatik.hu-berlin.de
Questions?