Exercise in Information Retrieval (Master), Summer term 2016
Ioannis Karatassis (LE 324)
Office hours: by appointment
karatassis@is.inf.uni-due.de

Exercise sheet 4 No submission

Exercise notes: By presenting (enough) exercise tasks, it is possible to receive bonus points for the (oral) exam. Those tasks are marked with a (*). You should prepare some slides for the task you want to present. We expect students to participate at discussions and to contribute input – in the form of presentations or questions – to the exercise that allows a good discussion.

http://www.is.inf.uni-due.de/courses/irmai_ss16

Presentation date: 6/23/2016

Task 1: Document features for Learning to Rank

Learning to Rank (LTR) is a means for predicting the ranking or retrieval status value (RSV) through machine learning methods. In this context, features of documents and queries play a significant role.

- What is the idea behind LTR methods and how do they work in principle?
- What features of documents and queries can be used in combination with LTR? Provide several examples by stating term based as well as other specific features for both the following use cases:
  - Web search (e.g., Google, Bing, or Yandex)
  - Product search (e.g., Amazon or Ebay)

Additionally, think of a third use case and provide useful features that can be used to perform LTR.
Task 2: Learning to Rank as classification problem

(a) Based on the discovered features in task 1), your task is now to formulate LTR as classification problem by first providing a complete table with sample training data for the scoring approach\(^1\). The data consists of the features and relevance judgements.

Example training data for the scoring approach:

<table>
<thead>
<tr>
<th></th>
<th>q</th>
<th>m₁</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>a</td>
<td>0.4</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td>b</td>
<td>0.634</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td>c</td>
<td>0.122</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

(b) Now, transform the table in (a) into sample training data for the ranking approach\(^2\).

(c) Which problems and challenges can arise with LTR in practice? Think of advantages and disadvantages.

---

\(^1\)RSV is being predicted

\(^2\)Order of documents is being predicted (ordinal classification)