CS 695: Topics in Information Retrieval

Instructor: Johan Bollen
Office: Education 228-3
Email: jbollen@cs.odu.edu
Homepage: www.cs.odu.edu/~jbollen
Office hours: Wednesday 2-5 PM
Class: ED0226, R 7:10-10, 3hr

Instructor Contact

The instructor keeps an open house policy. Students are allowed to consult at any time but no guarantees concerning availability or presence can be made outside office hours. Appointments can be made via e-mail or after class.

Students are encouraged to contact the instructor via e-mail but are asked to maintain generally accepted standards of e-mail etiquette. This means: proper salutations, no shouting (capitals), no demands or orders, no remarks or comments of a personal nature and the student’s full name and student ID. All e-mails that do not conform to the mentioned requirements will be ignored. Attachments of any nature and HTML code in the e-mail main body are strongly discouraged.

Course Objective

The domain of Information Retrieval is concerned with the extraction of relevant information from large collections of documents. It has applications to proprietary retrieval systems as well as the WWW, Digital Libraries and commercial recommendation systems. This course will aim to provide students with an overview of the main principles and methods underlying the domain of Information Retrieval. A number of advanced topics will be covered to address more recent developments in IR such as collaborative filtering and Latent Semantic Indexing. Students will furthermore acquire practical experience in the construction of IR systems by a series of projects.

Course Structure

This course is divided roughly into two parts. The first part will focus on the general principles underlying modern information retrieval systems including techniques for text processing and storage and retrieval. The second part will focus on advanced topics such as recommender system, WWW analysis, Latent Semantic indexing, etc.

Texts

Required:


Recommended:

- Michael W. Berry and Murray Browne (1999), Understanding search engines: mathematical modeling and text retrieval, Society for Industrial and Applied Mathematics.
- Dale Dougherty and Arnold Robbins(1997), Sed and awk, O’Reilly
Slides: copies of class slides may be provided. However, these slides do not substitute the student’s class notes or any other material: they serve merely as an educational aid. The instructor will not accept responsibility for failed study attempts using the provided class slides alone or a student’s failure to make adequate notes. The student is required to make class notes and is advised to apply all available materials (textbooks, notes, slides, etc) to their study.

Evaluation

The following distribution of scores will be used to determine your final grade for this class:

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midterm exam</td>
<td>30%</td>
</tr>
<tr>
<td>Final Exam</td>
<td>30%</td>
</tr>
<tr>
<td>Projects</td>
<td>40%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

All scores except final grade will be communicated on a numerical scale. The midterm exam may be on a 30 point scale, as well as the final exam. All scores will be added at the end of the course, taking into account their corresponding weight, and transformed into a final grade.

Late or inadequate submission

Late or inadequate project submission and presentations, i.e. submitted contrary to the stated requirements, will be failed. Scores will correspond to merit and quality of the delivered work, not necessarily to the student’s effort or desire.

Honor Code

It is a violation of the ODU Honor code to copy files, ideas, other people’s work or to collaborate on individual assignments in order to complete homework and assignments. It is not a violation to discuss a program with your fellow students or to give or receive advice about a project, but original work is required. Copies, or slightly altered copies, of available materials (e.g. WWW) and other students’ work will be regarded as inadequate submissions and will be failed. Serious offenses will be reported to the Honor Code Council.

Class Interaction

The instructor welcomes lively interaction and appreciates student creativity. Please do not hesitate to ask questions or make comments in class. Do not be intimidated by the difficulty of a specific subject. Please make contributions and speak your mind.

Disclaimer

The instructor reserves the right to alter the number and nature of assignments, quizzes, programming projects, the grade distribution, class presentation and content, and the required development and compilation tools to optimize learning outcomes and course logistics.

| I have read, understand and agree with the terms and conditions stated in this syllabus. |
|---------------------------------|------------------------------------------------|
| Student name:                   | Student ID:                                     |
| Signature                       |                                                |