COMPSCI 444: Introduction to Text Retrieval and Its Applications in Biomedicine

COMPSCI 744: Text Retrieval and Its Applications in Biomedicine

5:30 PM-6:50 PM  T  09/04-12/13  PHY 146

First day: Tuesday September 4, 2018

Note: this course will be taught blended with recorded primary lectures and in class exams and discussion.

Overview: Biomedical text and data has historically been fast growing and required expert knowledge to access effectively. The rise of the internet has only intensified this growth, challenging biomedical researchers and others to keep pace with this ever-increasing quantity of information. This situation presents opportunities and challenges for information retrieval (IR). This course addresses text retrieval and classification applications in biomedicine. Specifically the course will cover advanced levels of biomedical indexing, query processing, IR algorithms, and document retrieval methods involving supervised and unsupervised machine learning.

Textbook:

Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze. Introduction to Information Retrieval. Cambridge University Press, 2008 (available online too).

Online resources:

McRoy will provide recorded lectures via D2L, as Adobe flash files. Students will also be required to download and install (free) software tools, Lucene and Weka.

The authors’ webpage includes entire text book, in multiple formats:

There is also a long list of errata that appear in various printings of the book:


**Course Objectives:**

At the conclusion of this course, students will:

1. Understand why IR is important, especially in biomedicine.
2. Understand IR algorithms and techniques for indexing, query processing, and document classification and their applications in biomedicine.
3. Be able to use IR algorithms and software tools to solve problems.

**Grading:**

Undergrads:

30% Homework and online quizzes, 35% Midterm exam, 35% Final exam

Grads:

35% Homework and online quizzes, 25% Midterm 40% Software Project: The project may be done individually or in a team of 2-3 people. The final project will include a software system, a 4-6 page written project report, and a recorded oral presentation, including a demo of the working software. The report should describe the problem, the approach, and evaluation and should cite related work where appropriate.

**Tentative Timeline:**

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Resource Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Course Overview</td>
<td>L1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L1</td>
</tr>
<tr>
<td>Week 2</td>
<td>Introduction</td>
<td>L2</td>
</tr>
<tr>
<td></td>
<td>Indexing</td>
<td>L3_4</td>
</tr>
<tr>
<td>Week 3</td>
<td>Indexing (cont)</td>
<td>L3_4</td>
</tr>
<tr>
<td></td>
<td>Query processing</td>
<td>L5_7</td>
</tr>
<tr>
<td>Week 4</td>
<td>Query processing (cont)</td>
<td>L5_7</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Week 5</td>
<td>Usability and Query Expansion</td>
<td>L8_9</td>
</tr>
<tr>
<td>Week 6</td>
<td>Ranking, Scoring and term weighting Vector space models</td>
<td>L10 L11</td>
</tr>
<tr>
<td>Week 7</td>
<td>Evaluation <strong>MIDTERM EXAM</strong> <strong>Introduction to term projects</strong></td>
<td>L12_13</td>
</tr>
<tr>
<td>Week 8</td>
<td>Text classification</td>
<td></td>
</tr>
<tr>
<td>Week 9-10</td>
<td>Text classification (cont) XML retrieval</td>
<td></td>
</tr>
<tr>
<td>Week 11</td>
<td>Web searching</td>
<td></td>
</tr>
<tr>
<td>Week 12</td>
<td>Web searching (cont)</td>
<td></td>
</tr>
<tr>
<td>Week 13 -14</td>
<td>Project presentations &amp; Q/A on final</td>
<td></td>
</tr>
<tr>
<td>Final Exam</td>
<td>As published</td>
<td></td>
</tr>
</tbody>
</table>
Other Course Information:

- Students with Disabilities: Students who have special needs requiring special accommodations please contact the instructor. The Student Accessibility Center (Mitchell 112, 414-229-5822) is also an excellent resource and staff is available to discuss concerns.

- Religious Observances: Students will be allowed to complete examinations or other requirements that are missed because of a religious observance if prearranged with the instructor.

- Academic Misconduct: The University has a responsibility to promote academic honesty and integrity and to develop procedures to deal effectively with instances of academic dishonesty. Students are responsible for the honest completion and representation of their work, for the appropriate citation of sources, and for respect of other's academic endeavors. University policy prohibits and punishes misconduct, including any act by which a student seeks to claim credit for the work or efforts of another without authorization or citation (plagiarism), forges or falsifies documents, falsely represents his or her academic performance (cheating), or assists other students in any of these acts. Students who violate academic standards as set forth in UWS Chapter 14 and UWM Faculty Document 1686 will be confronted and must accept the consequences for their actions. Students who engage in academic misconduct are subject to a range of sanctions including but not limited to: a failing grade on an assignment or test, a failing grade in the course, and expulsion from the university.

- Sexual Harassment: Sexual harassment is reprehensible and will not be tolerated by the University. It subverts the mission of the University and threatens the careers, educational experience, and well being of the students, faculty, and staff. The University will not tolerate behavior between or among members of the University community, which creates an unacceptable environment.