Information Retrieval and Web Search – Fall 2022
COMP 7/8130
Vasile Rus

Contact Information:

<table>
<thead>
<tr>
<th>Office</th>
<th>Department Office</th>
</tr>
</thead>
<tbody>
<tr>
<td>323 Dunn Hall</td>
<td>375 Dunn Hall</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phone</th>
<th>Department Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>678-5259</td>
<td>678-5465</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E-mail</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:vrus@memphis.edu">vrus@memphis.edu</a></td>
</tr>
</tbody>
</table>

Office Hours (in DH323):

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:00PM-2:00PM</td>
<td>1:00PM-2:00PM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Also by Appointment

Course Description:

COMP 7130-8130. (from the University catalog).

Advanced current research topics in database and information management, with emphasis on nontraditional data and applications. PREREQUISITE: COMP 7116 or permission of instructor.

Why this course?

Today, one of the major problems is not lack of information but too much information. We need intelligent ways to organize the vast amount of information available at our fingertips and to effectively search for what we want. We are faced with an Information Overload problem. The Information Retrieval and Web Search class presents the major challenges posed by this problem and solutions to these challenges.

The course introduces computational techniques to search for information in static collections of documents and in dynamic collections, e.g. the web. Students will be exposed to text processing algorithms, classical information retrieval models (e.g., Boolean and Vectorial models), and web search techniques. The course is closely related to the Natural Language Processing course.

Resources:

See the class website: [http://www.cs.memphis.edu/~vrus/teaching/ir-websearch/](http://www.cs.memphis.edu/~vrus/teaching/ir-websearch/)
Required Text

R. Baeza-Yates and Berthier Ribeiro-Neto: Modern Information Retrieval

Recommended Texts

B. Frakes and R. Baeza-Yates Information Retrieval: Data Structures and Algorithms
C. Manning, P. Raghavan, and H. Schutze: Introduction to Information Retrieval

Other Resources:

See the class website: http://www.cs.memphis.edu/~vrus/teaching/ir-websearch/

Evaluation:

The University policy requires to email a grade to a student's U of M email address only.

Final Grades:
Homework 45%, Midterm 20%, Quiz #1, #2, #3 - 10% each, Participation & Presentations 5% PhD Students are required to make a presentation as part of the Homework grade. The presentation consists of power point slides presented in front of the class for 20-30 minutes and an accompanied report of 3-5 pages (single spaced).

Grading Scale:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Letter Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>90-100+</td>
<td>A</td>
</tr>
<tr>
<td>80-89</td>
<td>B</td>
</tr>
<tr>
<td>70-79</td>
<td>C</td>
</tr>
<tr>
<td>60-69</td>
<td>D</td>
</tr>
<tr>
<td>0-59</td>
<td>F</td>
</tr>
</tbody>
</table>

2.5 above or below the cut-off will earn you a + or – in front of your grade.
For example: 89 has a letter equivalent of B+
Exception: A- is for 90-91, A for scores ranging from 92 to 97, anything above 98 leads to A+

Course Policies:

Attendance
Students are strongly encouraged to attend all lectures. Active participation to class discussions counts toward your final grade.
Late Policy
Students will have on average one-two weeks from the date the work is assigned. Late submissions are not accepted. In exceptional cases you may have a 48-hour grace period at the cost of 50% of the grade (Students must ask for it before the due date).

Testing Policy
Usually exams are closed books. There are no make-up exams. Any code developed as part of the class work must follow the coding-style guidelines described on the web site. The coding-style will be strictly enforced.

Final Exam University General Rules and Guidelines

- Attendance is compulsory.
- No exam period may be changed without the written permission of the department chair and the college dean.

Plagiarism/Cheating Policy:

_plagiarism or cheating_ behavior in any form is unethical and detrimental to proper education and will not be tolerated. All work submitted by a student (projects, programming assignments, lab assignments, quizzes, tests, etc.) is expected to be a student's own work. The plagiarism is incurred when any part of anybody else's work is passed as your own (no proper credit is listed to the sources in your own work) so the reader is led to believe it is therefore your own effort. Students are allowed and encouraged to discuss with each other and look up resources in the literature (including the internet) on their assignments, but appropriate references must be included for the materials consulted, and appropriate citations made when the material is taken verbatim.

If plagiarism or cheating occurs, the student will receive a failing grade on the assignment and (at the instructor’s discretion) a failing grade in the course. The course instructor may also decide to forward the incident to the University Judicial Affairs Office for further disciplinary action. For further information on U of M code of student conduct and academic discipline procedures, please refer to:
http://www.people.memphis.edu/~jaffairs/
Course Syllabus (tentative)

Week 1: Introduction to IR and Web Search
Week 2: Introduction to Perl
Week 3: Classic IR: Boolean and Vectorial Models
Week 4: More on IR Models
Week 5: Evaluation of IR
Week 6: Query Operations and Languages
Week 7: Text Properties, Text Operations
Week 8: FALL BREAK, Indexing, and Searching
Week 9: MIDTERM, Web Search
Week 10: Web Search
Week 11: Text Categorization
Week 12: Text Clustering
Week 13: Question Answering
Week 14: Advanced IR Models, THANKSGIVING
Week 15: Project Presentations
Week 16: Final Exam