CS734/834: Introduction to Information Retrieval (Spring 2019)

Course Information
- Credit: 3
- Prerequisites: programming in Java/Python, algorithm, basic statistics, Linux system
- First class: 1/14/2019
- Last class: 4/29/2019
- Lecture time: T/R 9:30 am. - 10:45 am.
- Office Hours: Tuesday, 11 am. - noon, or by appointment
- Classroom: 2120 ECSB
- Textbook:
- Communication: All course-related issues should be posted on the Google group named *odu-cs834-s19*. Enrolled students will receive invitations to join this group. Class-materials including slides will be available at a box.com shared folder. Homework solutions will be uploaded to the same shared folder at box.com. Enrolled students will receive invitations to access this folder.

Description
This class will explore the theory and engineering of information retrieval in the context of developing web-based search engines. The course will explore issues related to crawling, ranking, query processing, retrieval models, evaluation, clustering, machine learning, and other aspects related to building search engines. The course will also cover recently established ranking algorithms that incorporate semantic similarities, machine learning, and neural network methods, such as learning to rank and neural information retrieval. The class will feature several hands-on development and coding using tools such as Google Custom Search, ElasticSearch, as well as a theoretical exploration of the existing literature on these topics. An industrial speaker will also be invited to give a talk on contemporary search engine and related topics (depending on availability). Students must be comfortable with self-directed learning appropriate for an advanced graduate class.

Grading Scale
1. Attendance: 10%
2. Homework: 30%
3. Advanced topic presentation: 20%
4. Project: 40%

Grading Chart

<table>
<thead>
<tr>
<th>Score Grades</th>
<th>Letter Grades</th>
</tr>
</thead>
<tbody>
<tr>
<td>94-100</td>
<td>A</td>
</tr>
<tr>
<td>90-93.99</td>
<td>A-</td>
</tr>
<tr>
<td>87-89.99</td>
<td>B+</td>
</tr>
<tr>
<td>84-86.99</td>
<td>B</td>
</tr>
<tr>
<td>80-83.99</td>
<td>B-</td>
</tr>
</tbody>
</table>
Attendance Policy
Attendance is required. One no-excuse absence causes a deduction of 1% on attendance until all points are deducted in this item. If more than 11 absence is observed, the student automatically gets “F” for this course.

In case of absence due to legitimate reasons, including but not limited to sickness, University-approved curricular and extracurricular activities (such as athletic contests), career interviews, the death of family members, students should be prepared to provide documentation before classes. Opportunities for makeup classes are available during office hours.

Late submission policy
Homework assignments and reports are due at exactly at 9 am. on the specified dates. Any submission after this deadline will be marked “late” in the instructor’s spreadsheet.
Each student has one chance to submit an overdue assignment for whatever reason within a grace period of 24 hours after deadlines without penalty. In these cases, students MUST provide a memo in advance to provide a reasonable explanation.
Submissions after deadlines or beyond grace period are counted 50% of the real score, e.g., if a student earns 80 points for his homework assignment, only 40 points are counted.

Academic integrity
Individual assignments must be completed independently. Students are strongly encouraged to form study groups and to learn from the peers. However, discussion on final proposal writing and presentation in a study group should be limited to general approaches to solutions. Specific answers should never be discussed. ODU’s policy regarding Academic integrity must be followed. The ODU requires students and faculties to respect honor code and academic integrity.

- **Cheating**: Using unauthorized assistance, materials, study aids, or other information in any academic exercise turned in for grading (Examples of unauthorized assistance include, but are not limited to using unapproved resources or assistance to complete assignments, papers, projects, quizzes or exams, collaborating in violation of a faculty member’s instructions, and submitting the same, or substantially the same, papers to more than one course for academic credits without first obtaining the approval of faculty).

- **Plagiarism**: Using someone else's language, ideas, or other original material without acknowledging its source in any academic exercise. 4 Examples of plagiarism include, but are not limited to the following: submitting a research paper obtained from a commercial research service, the Internet, or from another student as if it were original work; or making simple changes to borrowed materials while leaving the organization, content, or phraseology intact. Plagiarism also occurs in a group project if one or more of the members of the group does none of the group's work and participates in none of the group’s activities, but attempts to take credit for the work of the group.

- **Fabrication**: Inventing, altering or falsifying any data, citation or information in any academic exercise. Examples of fabrication include, but are not limited to, the following: citation of a primary source which the student actually obtained from a secondary source; or invention or alteration of experimental data without appropriate documentation (such as statistical outliers).

- **Facilitation**: Helping another student commit, or attempt to commit, any Academic Integrity violation, or failure to report suspected Academic Integrity violations to a faculty member. An example of facilitation may include circulating course materials when the faculty member has not explicitly authorized their use.

Copyrights
All course materials students receive or to which students have online access are protected by copyright. FOR COURSES IN WHICH THEY HAVE PREVIOUSLY BEEN OR ARE CURRENTLY ENROLLED, students may use course materials and make copies for their own use as needed, but unauthorized distribution and/or uploading of materials without the instructor's express permission is strictly prohibited. For example, students may not post details of course assignments, projects, or tests at online Forums, Bulletin Boards, Homework sites, etc., soliciting help.

When writing a paper or a computer program, if you use someone else's ideas, you must cite your sources appropriately. Failure to appropriately cite any such “found code” will be taken as evidence of plagiarism.

Disability Accommodations
In order to receive consideration for reasonable accommodations, you must contact the appropriate disability services office at the campus where you are officially enrolled and provide documentation. If the documentation supports your request for reasonable accommodations, your campus's disability services office will provide you with an accommodation letter. Please share this letter with your instructors and discuss the accommodations with them as early in your courses as possible. The detail of disability accommodations is documented in ODU policy #4500.

Discrimination and Harassment
The university is committed to equal access to programs, facilities, admission and employment for all persons. It is the policy of the university to maintain an environment free of harassment and free of discrimination against any person because of age, race, color, ancestry, national origin, religion, creed,
disability, gender, perceived gender, gender identity, genetic information or political ideas. Discriminatory conduct and harassment, as well as sexual misconduct and relationship violence, violates the dignity of individuals, impedes the realization of the university’s educational mission, and will not be tolerated.

Gender-based sexual harassment, including sexual violence, are forms of gender discrimination in that they deny or limit an individual’s ability to participate in or benefit from University programs or activities. These policies shall not be construed to restrict academic freedom at the university, nor shall they be construed to restrict constitutionally protected expression. The policy is coded in University Policy #1005.

### Course Schedule

<table>
<thead>
<tr>
<th>Course#</th>
<th>Dates</th>
<th>Topics</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tuesday, January 15, 2019</td>
<td>Course Introduction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thursday, January 17, 2019</td>
<td>Introduction to IR and Search Engines</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Tuesday, January 22, 2019</td>
<td>Complexity and scalability of search</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Thursday, January 24, 2019</td>
<td>IR evaluation</td>
<td>Assignment 1</td>
</tr>
<tr>
<td>4</td>
<td>Tuesday, January 29, 2019</td>
<td>No class (instructor traveling)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Thursday, January 31, 2019</td>
<td>No class (instructor traveling)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Tuesday, February 5, 2019</td>
<td>student presentation</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Thursday, February 7, 2019</td>
<td>student presentation</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Tuesday, February 12, 2019</td>
<td>Specialty search</td>
<td>Assignment 1 due</td>
</tr>
<tr>
<td>9</td>
<td>Thursday, February 14, 2019</td>
<td>Web Crawling and Properties of text</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Tuesday, February 19, 2019</td>
<td>Text Processing</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Thursday, February 21, 2019</td>
<td>Query models</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Tuesday, February 26, 2019</td>
<td>Vector models</td>
<td>Assignment 2</td>
</tr>
<tr>
<td>13</td>
<td>Thursday, February 28, 2019</td>
<td>Indexing-1 and ElasticSearch</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Tuesday, March 5, 2019</td>
<td>Project proposal presentations</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Thursday, March 7, 2019</td>
<td>Indexing-2 and Google custom search</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tuesday, March 12, 2019</td>
<td>Spring break, no class</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thursday, March 14, 2019</td>
<td>Spring break, no class</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Tuesday, March 19, 2019</td>
<td>Web search (1)</td>
<td>Assignment 2 due</td>
</tr>
<tr>
<td>17</td>
<td>Thursday, March 21, 2019</td>
<td>Invited talk (Kyle Williams)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Tuesday, March 26, 2019</td>
<td>Web search (1)</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Thursday, March 28, 2019</td>
<td>Student presentation</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Tuesday, April 2, 2019</td>
<td>Student presentation</td>
<td>Assignment 3</td>
</tr>
<tr>
<td>21</td>
<td>Thursday, April 4, 2019</td>
<td>Link analysis</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Tuesday, April 9, 2019</td>
<td>Project status report</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Thursday, April 11, 2019</td>
<td>XML and semantic web</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Tuesday, April 16, 2019</td>
<td>Web 3.0 and Web 4.0</td>
<td>Assignment 3 due</td>
</tr>
<tr>
<td>25</td>
<td>Thursday, April 18, 2019</td>
<td>Project presentation</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Tuesday, April 23, 2019</td>
<td>Project presentation</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Thursday, April 25, 2019</td>
<td>Project report due at noon</td>
<td></td>
</tr>
</tbody>
</table>