<table>
<thead>
<tr>
<th>Module Code</th>
<th>CS7IS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module Name</td>
<td>INFORMATION RETRIEVAL AND WEB SEARCH</td>
</tr>
<tr>
<td>ECTS Weighting</td>
<td>5 ECTS</td>
</tr>
<tr>
<td>Semester taught</td>
<td>Semester 2</td>
</tr>
<tr>
<td>Module Coordinator/s</td>
<td>Owen Conlan and Joeran Beel</td>
</tr>
</tbody>
</table>

### Module Learning Outcomes

On successful completion of this module, students will be able to:

- **LO1.** Explain the process of content indexing in information retrieval including stop word removal, conflations (stemming, string-comparison), and the language dependency of these methods.
- **LO2.** Demonstrate an understanding of the importance and application of data structures in efficient information retrieval, in particular inverted file structures.
- **LO3.** Have knowledge of the theoretical basis and operation of standard algorithms for ranked information retrieval, including the term weighting and ranking models e.g. tf-idf weighting, vector-space model, probabilistic model, language modelling.
- **LO4.** Describe the process of relevance feedback for improved ranking in information retrieval, and apply standard relevance feedback algorithms.
- **LO5.** Understand the importance of evaluation in development of search engines, and the application of standard evaluation metrics such as precision and recall and test collections in measuring effectiveness of information retrieval systems, both in terms of the system's performance and user satisfaction with the system.
- **LO6.** Appreciate the application and operation of search engines in diverse environments e.g. web search, audio-visual search, context-aware and mobile search, patent search, search in microblogs etc.
- **LO7.** Be able to begin to combine technologies relevant to search systems in novel ways to synthesise new information retrieval applications.

### Module Content

Specific topics addressed in this module include:

- Introduction to Web Search
- Boolean Retrieval
- Text Processing: Stopword Removal, Stemming, Spelling Correction...
- Index Construction and Compression
- Probabilistic Information Retrieval
- Computing Scores for Ranking: BM25, Vector Space Model, PageRank...
- Classification: Naive Bayes, kNN, decision boundaries
- Evaluation: Precision, Recall, F-score, NDCG...
- Link Analysis
- Web Crawling
- Question Answering
- Personalisation
Teaching and Learning Methods

Lectures, labs and self-directed study

### Assessment Details

<table>
<thead>
<tr>
<th>Assessment Component</th>
<th>Brief Description</th>
<th>Learning Outcomes Addressed</th>
<th>% of total</th>
<th>Week set</th>
<th>Week due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examination</td>
<td>2 Hour Exam</td>
<td>50%</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Assignments</td>
<td>Major Assignments</td>
<td>50%</td>
<td>TBC</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

### Reassessment Details

100% based on Examination

### Contact Hours and Indicative Student Workload

<table>
<thead>
<tr>
<th>Contact Hours (scheduled hours per student over full module), broken down by:</th>
<th>22 hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture + labs</td>
<td>22 hours</td>
</tr>
<tr>
<td>Independent study (outside scheduled contact hours), broken down by:</td>
<td>103 hours</td>
</tr>
<tr>
<td>Self-directed study + completion of assignments</td>
<td>103 hours</td>
</tr>
<tr>
<td>Total Hours</td>
<td>125 hours</td>
</tr>
</tbody>
</table>

### Recommended Reading List


### Module Pre-requisites

Prerequisite modules: n/a

Other/alternative non-module prerequisites: n/a

### Module Co-requisites

mymodule.tcd.ie

### Module Website

09/02/2020 by Owen Conlan and Joeran Beel

---

2 TEP Guidelines on Workload and Assessment