270657 - IR - Information Retrieval

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science
Academic year: 2019
Degree: MASTER’S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Teaching unit Optional)
ECTS credits: 6
Teaching languages: English

Prior skills

Those assumed for admission to MIRI plus those provided by the common learning phase.

Teaching methodology

Sessions of theory + problems of 3 sessions per week. The 2 hours of each session are theoretical expositions, and the third one is devoted to joint exercise solving. For each session, the student will have to deliver solutions to a few problems proposed but not solved in the previous session.

Laboratory sessions of 1 hour per week. For many of the sessions, the student will have to deliver a report of the work done and obtained results after about two weeks.

The working of each type of session is described in the "Activities" session.

Furthermore, at the end of the course each student must present to instructors and fellow students a scientific paper related to the course topic, in the format of a conference presentation. Near week 8 of the course, a list of papers will be made public, from which each student can choose one, or alternatively propose a paper of his/her choice, to be approved by the instructors. The date and time range for the presentations will be announced with at least 2 months time, and the schedule within the chosen day at least 1 week time.

Learning objectives of the subject

Study load

<table>
<thead>
<tr>
<th>Total learning time: 150h</th>
<th>Hours large group: 24h</th>
<th>16.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours medium group: 12h</td>
<td>8.00%</td>
</tr>
<tr>
<td></td>
<td>Hours small group: 12h</td>
<td>8.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 6h</td>
<td>4.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 96h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
## Introduction

**Degree competences to which the content contributes:**

**Description:**
Need of search and analysis techniques of massive information. Search and analysis vs. databases. Information retrieval process. Preprocessing and lexical analysis.

## Models of information retrieval

**Degree competences to which the content contributes:**

**Description:**

## Implementation: Indexing and searching

**Degree competences to which the content contributes:**

**Description:**
Inverse and signature files. Index compression. Example: Efficient implementation of the rule of the cosine measure with tf-idf. Example: Lucene.

## Evaluation in information retrieval

**Degree competences to which the content contributes:**

**Description:**
Recall and precision. Other performance measures. Reference collections. Relevance feedback and query expansion.

## Web search

**Degree competences to which the content contributes:**

**Description:**

## Architecture of massive information processing systems

**Degree competences to which the content contributes:**
### Description:

### Network analysis

**Degree competences to which the content contributes:**

**Description:**
Descriptive parameters and characteristics of networks: degree, diameter, small-world networks, among others. Algorithms on networks: clustering, community detection and detection of influential nodes, reputation, among others.

### Information Systems based on massive information analysis. Combination with other technologies.

**Degree competences to which the content contributes:**

**Description:**
## Planning of activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Hours</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theoretical development of topics 1 to 8 of</strong></td>
<td>52h</td>
<td>The student will attend the instructor's presentation and actively participate in the initial discussion of the challenge to be solved in that session.</td>
</tr>
<tr>
<td>the course</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Exercises on topics 1 to 8 of the course</strong></td>
<td>39h</td>
<td>In each session, the instructor proposes a number of exercises (say, 4 to 7) on the topic just covered in theory. Next, a few of the problems (say, 3) are solved jointly. Students must solve the rest of the exercises and deliver them by the start of next session. A part of the session is devoted to discussing the possible questions that may have appeared while solving the problems pending from the last session.</td>
</tr>
<tr>
<td><strong>Laboratory work on topics 1 to 8</strong></td>
<td>26h</td>
<td>The teacher will describe a practical work to be carried out related with the topics most recently covered. This may be a data analysis task, the implementation of an algorithm seen in class, or proposing a solution for an Information Retrieval scenario. The student completes the work as much as possible in class, although occasionally some additional time may be necessary. In many cases the student will have to produce a report on the work done and results obtained, to be delivered within some clearly stated deadline (say, 2 weeks).</td>
</tr>
<tr>
<td><strong>Final exam</strong></td>
<td>18h</td>
<td>Final exam on the contents of the whole course</td>
</tr>
</tbody>
</table>

### Theory classes
- Theory classes: 26h
- Practical classes: 0h
- Laboratory classes: 0h
- Guided activities: 0h
- Self study: 26h

### Practical classes
- Theory classes: 0h
- Practical classes: 13h
- Laboratory classes: 0h
- Guided activities: 0h
- Self study: 26h

### Laboratory classes
- Theory classes: 0h
- Practical classes: 0h
- Laboratory classes: 13h
- Guided activities: 0h
- Self study: 13h

### Guided activities
- Theory classes: 0h
- Practical classes: 0h
- Laboratory classes: 0h
- Guided activities: 0h
- Self study: 13h

### Self study
- Theory classes: 26h
- Practical classes: 13h
- Laboratory classes: 13h
- Guided activities: 18h
- Self study: 18h
270657 - IR - Information Retrieval

**Study and presentation of a scientific paper**

<table>
<thead>
<tr>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 0h</td>
</tr>
<tr>
<td>Practical classes: 0h</td>
</tr>
<tr>
<td>Laboratory classes: 0h</td>
</tr>
<tr>
<td>Guided activities: 3h</td>
</tr>
<tr>
<td>Self study: 10h</td>
</tr>
</tbody>
</table>

**Description:**
Study and presentation of a scientific paper related to the course topic

**Qualification system**

Define:

- NF as the grade of the final exam
- NE the grade of exercise assignments
- NL the grade of lab reports
- NA the grade from the presentation of a scientific article

(all in the range 0..10).

Then the final course grade is 0.3*NF + 0.25*NL + 0.25*NE + 0.2*NA.

**Bibliography**

**Basic:**


**Others resources:**

- Hyperlink
  