270657 - IR - Information Retrieval

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science
Academic year: 2018
Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Teaching unit Optional)
ECTS credits: 6
Teaching languages: Catalan

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

Degree competences to which the subject contributes

Basic:
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

Specific:
CEC1. Ability to apply scientific methodologies in the study and analysis of phenomena and systems in any field of Information Technology as well as in the conception, design and implementation of innovative and original computing solutions.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CEC3. Ability to apply innovative solutions and make progress in the knowledge that exploit the new paradigms of Informatics, particularly in distributed environments.

Genercal:
CG1. Capability to apply the scientific method to study and analyse of phenomena and systems in any area of Computer Science, and in the conception, design and implementation of innovative and original solutions.
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CG5. Capability to apply innovative solutions and make progress in the knowledge to exploit the new paradigms of computing, particularly in distributed environments.

Transversal:
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR5. APPROPRIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
270657 - IR - Information Retrieval

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>Theory classes: 24h</th>
<th>16.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 12h</td>
<td>8.00%</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 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:**
### 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

| Theoretical development of topics 1 to 8 of the course | Hours: 52h  
Theory classes: 26h  
Practical classes: 0h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 26h |
|-----------------------------------------------------|--------------------------------------------------|
| Description:  
The student will attend the instructor's presentation and actively participate in the initial discussion of the challenge to be solved in that session. |

| Exercises on topics 1 to 8 of the course | Hours: 39h  
Theory classes: 0h  
Practical classes: 13h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 26h |
|-----------------------------------------|--------------------------------------------------|
| Description:  
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. |

| Laboratory work on topics 1 to 8 | Hours: 26h  
Theory classes: 0h  
Practical classes: 0h  
Laboratory classes: 13h  
Guided activities: 0h  
Self study: 13h |
|----------------------------------|--------------------------------------------------|
| Description:  
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). |

| Final exam | Hours: 18h  
Guided activities: 3h  
Self study: 15h |
|----------------|--------------------------------------------------|
| Description:  
Final exam on the contents of the whole course |
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:**


Manning, Christopher D; Raghavan, Prabhakar; Schütze, Hinrich. Introduction to information retrieval. Cambridge University Press, 2008. ISBN 9780521865715.


Russell, Matthew A. Mining the social web [Recurs electrònic]: [data mining Facebook, Twitter, LinkedIn, Google+, Github, and more]. 2nd ed. O'Reilly, 2013. ISBN 9781449368227.


**Others resources:**

Hyperlink

http://www.cs.upc.edu/~IR-MIRI/