Information access and retrieval

Academic Year: (2020 / 2021) **Review date: 29-10-2020**

Department assigned to the subject: Computer Science and Engineering Department

Coordinating teacher: MORATO LARA, JORGE LUIS

Type: Compulsory ECTS Credits : 6.0

Year : 4 Semester : 2

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

- Files and Data bases (Bachelor in Informatics Engineering, 2nd Course, Semester 2nd, Compulsory)
- Object oriented programming (Bachelor in Informatics Engineering, 1st Course, 2nd Semester, Compulsory)

OBJECTIVES

(Competences related with the ABET program are displayed in parentheses)
General Competences:
- Systematic acquisition of theoretical concepts (PO: i)
- Ability to organize and communicate results (PO:g)
- Ability to apply theoretical concepts in real-world scenarios (PO:a, b, c, e, k)
- Teamwork (PO:d)
- Problem solving in multidisciplinary contexts (PO:c)

Specific Competences:
- Cognitive (PO:a, c, j, k)

1. Retrieval models
2. Natural Language Processing Techniques
3. Systems to formalize, synthesize, and structure information
4. Traceability systems
5. Ability to show results in an appropriate way
6. Improve retrieval and knowledge reuse systems in the Web and in Software Engineering

- Procedimental/Instrumental Competences (PO:a, b, c, e, j)

1. Design of Retrieval Systems
2. Design of natural language analyzers
3. Application of text mining techniques to improve the representation and sorting of results

- Attitudinal Competences (PO: c, e, i)

1. Concern for quality results
2. Capability to solve problems in an autonomous manner
3. Encouragement to independent research and acquisition of knowledge necessary to solve problems

DESCRIPTION OF CONTENTS: PROGRAMME

Description: Retrieval Models, Natural Language Processing, semantic analysis, metadata, linked data, information retrieval, positioning techniques, knowledge reuse, data mining
The course examines fundamental concepts about retrieval systems, introducing a variete of basic techniques. This includes the use of knowledge organization systems, positioning techniques, natural language processing techniques and resources, and evaluation by retrieval metrics.

Course content, 3 units:
Unit 1. Information retrival
- Lesson 1: Search basics in different web types: classic web, Semantic Web, Social Web, Data Web, Dark Web, Deep Web, question-answering web, and commercial web.
- Lesson 2: Search Engine Optimization (SEO/SEM)
- Lesson 3. Basic information retrieval models
- Lesson 4: Access, acquisition and cleansing of semantic web data and bigdata
- Lesson 5. Crawlers, scrapers and search engine arquitecture

Unit 2. Retrieval evaluation

- Lesson 6. Evaluation metrics for information retrieval systems

Unit 3. Advanced techniques for information retrieval systems
- Lesson 7. Natural Language Processing (NLP)
- Lesson 8. Information extraction techniques (IE)
- Lesson 9. Relevance feedback and query expansion

LEARNING ACTIVITIES AND METHODOLOGY

Theoretical lectures: 1.5 ECTS. To achieve the specific cognitive competences of the course (PO: a, c, j, k)
Practical lectures: 1.5 ECTS. To develop the attitudinal and specific competences as well as most of the general ones, such as collaborative teamwork, skills to apply theoretical concepts, design planning, information organization, analysis, and abstraction.
Students must design and develop an information retrieval system. Practical exercises deal with web positioning algorithms and retrieval metrics and technologies (PO: a, b, c, d, e, i, j)
- Academic activities with the professor: 1 ECTS. Students must carry on collaborative work to evaluate their ability to apply theoretical concepts and meet the desired needs. (PO: a, b, c, d, e, g, k)
- Guided academic activities (absent teacher): 0.5 ECTS. Complementary homework and technical readings suggested by the professor (PO: j)
Exercises and examination: 1.5 ECTS. The goal is to complete the development of specific cognitive and procedural capacities. Exercises and results of the practices will be discussed in class. (PO: a, c, g)

The course includes two hours per week of one-on-one tutorial

ASSESSMENT SYSTEM

Exercises and examinations serve for both the learning and the valuation processes.
Assessment criteria:
- Class exercises: Quality, organization and correctness of a written report
- Collaborative Work: oral presentation, relevance judgments and the application's performance and reliability
- Web page: quality of texts in English and Spanish; web page ranking in the Google Search Engine; optimization factors applied; novelty and quality of the contents.

The evaluation system includes the assessment of guided academic activities and practical cases, with the following weights:
Exam: 40% (minimum mark 4.0) (PO: a, c, g)
Practical cases: 40% (PO: a, b, c, d, e, i, j)
Guided academic activities and presentation: 20% (PO: a, b, c, d, e, g, j, k)

% end-of-term-examination:	40
% of continuous assessment (assigments, laboratory, practicals…):	60

BASIC BIBLIOGRAPHY

- Aurelien Geron Hands-On Machine Learning with Scikit-Learn and TensorFlow, OReilly, 2017
- Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda Applied Text Analysis with Python, OReilly, 2018
- R. Baeza-Yates y B. Ribeiro-Neto Modern Information Retrieval: The Concepts and Technology behind Search (2nd edition), Addison Wesley, 2011
- Verborgh, R., De Wilde, M., & Sawant, A. Using OpenRefine: The essential OpenRefine guide that takes you from data analysis and error fixing to linking your dataset to the web, Packt Publishing, 2013

ADDITIONAL BIBLIOGRAPHY

- Anne Ahola Ward The SEO battlefield: winning strategies for search marketing programs, OReilly, 2017
- Dale R Handbook of Natural Language Processing, Marcel Dekker, 2000
- Dean Allemang, James Hendler Semantic Web for the Working Ontologists: Effective Modeling in RDFS and OWL, Elservier, 2011
- Gábor László Hajba Website Scraping with Python: Using BeautifulSoup, Google Books, 2018
- Ian H. Witten, Alistair Moffat and Timothy C. Bell Managing Gigabytes: compressing and indexing documents, Morgan Kauffman, 1999
- J. Urbano, M. Marrero, D. Martín y J. Morato Bringing Undergraduate Students Closer to a Real-World Information Retrieval Setting: Methodology and Resources, ACM SIGCSE ITiCSE, 2011
- Moens Marie-Francine Information Extraction: algorithms and prospects in a retrieval context (Chps.

1, 2 & 4), Springer, 2006

- Morato, J, Sánchez-Cuadrado, S, Moreno, V Moreiro JA Evolución de los factores de posicionamiento web y adaptación de las herramientas de optimización, Revista española de Documentación Científica, Vol 36, No 3, 2013

- Nadeau D. and Sekine S. A survey of named entity recognition and classification, Linguisticae Investigationes vol. 30 n.1, 2007

- Stuart Russell, Peter Norvig Artificial Intelligence: A Modern Approach, Pearson, 2018

BASIC ELECTRONIC RESOURCES

- . How to Sparql. My experiment.: http://rdf.myexperiment.org/howtosparql

- . Google. Search Engine Optimization (SEO) Starter Guide : https://support.google.com/webmasters/answer/7451184?hl=en

- . MOZ. The beginner's guide to SEO: https://moz.com/beginners-guide-to-seo

- . W3schools tutorials (HTML, CSS, JS): https://www.w3schools.com/

- Manning, Christopher D; Raghavan, Prabhakar; Schütze, Hinrich . Introduction to Information Retrieval. Cambridge University Press: https://nlp.stanford.edu/IR-book/information-retrieval-book.html