Web Analytics

Academic Year: (2022 / 2023) **Review date: 31-08-2021**

Department assigned to the subject: Telematic Engineering Department

Coordinating teacher: CUEVAS RUMIN, RUBEN

Type: Compulsory ECTS Credits : 6.0

Year : 4 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Programming (Course 1, Semester 1)
Data structures and Algorithms (Course 1, Semester 2)
Data Base (Course 2, Semester 1)
Web Applications (Course 3, Semester 1)

OBJECTIVES

1. Students should be able to demonstrate they have acquired and understood the knowledge associated to an area that starts from high school education and reach a level that although it is based on text books, it also includes aspects that include concepts coming from up-to-date knowledge in the referread area.
2. Students should be able to apply the acquired knowledge to their job in a professional way and should incorporate the required competences that can be shown through solid arguments and the resolution of problems within their area of study.
3. Ability to design solutions based on automatic knowledge within applications applied to specific domains such as: recommendation systems, natural language processing, the WEB or online social networks.
4. Ability to develop web and mobile applications and crawlers to collect data using them.
5. Ability to develop data visualization tools to communicate the results derived from data analysis.
6. Adequate knowledge and skills to analyze and synthesize basic problems related to engineering and data science, solve them and communicate them efficiently
7. Ability to solve problems with initiative, decision making, creativity, and to communicate and transmit knowledge, skills and abilities, understanding the ethical, social and professional responsibility of the data processing activity. Leadership capacity, innovation and entrepreneurial spirit
8. Ability to communicate knowledge orally and in writing to both specialised and non-specialised audiences
9. Students should have acquired advanced knowledge and demonstrated an understanding of the theoretical and practical aspects and working
methodology in the field of data science and engineering with a depth that reaches the forefront of knowledge
10. Be capable of applying their knowledge and problem-solving skills, through arguments or procedures developed and sustained by themselves, in
complex or professional and specialized work settings that require the use of creative and innovative ideas.
11. Have the ability to collect and interpret data and information on which to base their conclusions including, where appropriate and pertinent, reflection on issues of a social, scientific or ethical nature within their field of study

DESCRIPTION OF CONTENTS: PROGRAMME

1. Data Collection in the Web ecosystem:
   - 1.1 Scrapers, Crawlers
   - 1.2 APIs
2. Data Analytics in the web
   - 2.1 Graph Analysis: Centrality and Influence metrics
   - 2.2 Network structure:
     - 2.2.1 Type of networks (bipartite graph, small world, scale free)
     - 2.2.2 Clustering, Community Detection, K-core decomposition
3. Web data visualization
   - 3.1 Representation of web information.

3.2 Visualization tools.

4. Final Web Analytics Project

4.1 The project needs to include the three components presented above (Data Collection, Data Analytics and Data Visualization

LEARNING ACTIVITIES AND METHODOLOGY

The course will be based in the following activities:

- LECTURES: theoretical lessons that will introduce the main concepts of the course. Students participation to discuss the concepts and problems introduced in the lectures will be encouraged.

- LABS: practical lessons in which students will bring to practice the concepts introduced in lectures. Students will have to solve practical problems associated to web analytics.

- FINAL GROUP PROJECT: Students will be assigned a project that will be developed throughout the semester in groups of 2 oe 3 people. Students should propose their own project. In exceptional cases the professors may offer a list of projects to students. The responsible professor has to approve the student proposal. The project will include the following elements:

1- An initial definition of the goals of the project, technology used and expected results

2- Use of any of the data collection studied to retrieve information from some popular online service or social network.

3- Data analysis using up to date technological frameworks (for instance python, R, etc).

4- Results visualization.

The students will defend their project in a public exposition to the rest of students at the end of the semester.

There will be a number of lab classes that will be used to supervise the evolution of the project and to allow students progressing in its development.

OFFICE HOURS: The students will get access to meetings with professors every week individually or collectively in order to clarify theorical and/or practical concepts. In addition, these meetings are valid to access to a more detailed supervision of student projects.

ASSESSMENT SYSTEM

- Labs (50%): Each lab practice defines a number of milestones to be achieved. Each achieve milestone adds up to the mark of such lap practice. In the case all milestones are successfully achieved the maximum mark for the lab will be obtained.
There are a total of 4 labs. Each of the lab practices has a weight of 12,5%.

-Final Project (50%): The assessment will be based on the technical correctness of the project, the volume of data collected and processed, the complexity of the project, and the efficiency of the adopted solution.

To successfully pass this course, both parts (Labs and Final projects) need to be passed independently. This means, the minimum grade out of the 100% required is 20% for the labs and 20% for the final project.

% end-of-term-examination:	0
% of continuous assessment (assigments, laboratory, practicals…):	100

BASIC BIBLIOGRAPHY

- Albert-László Barabási Network Science, http://networksciencebook.com/.

- Christopher Olston, Marc Najork. Web Crawling, Now Publishers Inc, , 2010

- David Easley and Jon Kleinberg Networks, Crowds, and Markets, Cambridge University Press. https://www.cs.cornell.edu/home/kleinber/networks-book/, 2010

- Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman Mining of massive datasets, Cambridge University Press., 2014

- Leonard Richardson, Mike Amundsen, Sam Rub Restful Web APIs, O¿Reilly, 2013

- Ryan Mitchell Web Scraping with Python, O'Reilly Media, Inc., 2018

- Stanley Wasserman, Katherine Faust Social Network Analysis: Methods and Applications, Cambridge University Press., 1994

BASIC ELECTRONIC RESOURCES

- Altexsoft . Comparing API Architectural Styles: SOAP vs REST vs GraphQL vs RPC: https://www.altexsoft.com/blog/soap-vs-rest-vs-graphql-vs-rpc/
- Camile Siegel . What is an API?: https://apifriends.com/api-management/what-is-an-api/
- GraphQL Foundation . A query language for your API: https://graphql.org/
- Jure Leskovec . (Social and Information) Network Analysis: http://snap.stanford.edu/na09/resources.html
- Pedro Pregueiro . Python & APIs: A Winning Combo for Reading Public Data: https://realpython.com/python-api/
- Red Hat . What is an API?: https://www.redhat.com/en/topics/api/what-are-application-programming-interfaces
- Red Hat . API Security: https://www.redhat.com/en/topics/security/api-security