

COURSE DESCRIPTION – ACADEMIC YEAR 2018/2019

Course title	Web and Text Mining
Course code	73015
Scientific sector	INF/01
Degree	Master in Computational Data Science (LM-18)
Semester	2
Year	1
Credits	6
Modular	No

Total lecturing hours	40
Total lab hours	20
Attendance	Attendance to lectures and labs is not compulsory. Non-attending students should contact the lecturer at the start of the course to agree on the modalities of the independent study.
Prerequisites	Basic programming skills for data analytics and foundations on probability theory and statistics.
Course page	https://ole.unibz.it/

Specific educational objectives	<p>The course belongs to the type "caratterizzanti – discipline informatiche" in the curricula "Data Analytics" and "Data Management".</p> <p>In this course the students will develop an understanding of fundamental natural language processing techniques used for acquiring and representing knowledge from the Web.</p> <p>The students shall also get in contact with real applications of the techniques studied and get a feeling of how to apply the theoretical knowledge obtained.</p>
--	--

Lecturer	Ana Ozaki
Contact	Piazza Domenicani 3 , Room 3.04, ana.ozaki@unibz.it
Scientific sector of lecturer	INF/01
Teaching language	English
Office hours	Check the homepage of the lecturer
Lecturing Assistant (if any)	—
Contact LA	—
Office hours LA	—
List of topics	<ul style="list-style-type: none"> • Text Representation (lexical, syntactical and semantical levels) • Document summarization, clustering and categorization • Entity extraction • Link mining • Social network analysis • Sentiment analysis
Teaching format	Frontal lectures, exercises and projects in teams.

Learning outcomes	Applying knowledge and understanding:
--------------------------	---------------------------------------

	<ul style="list-style-type: none"> • D2.1 - Practical application and evaluation of tools and techniques in the field of data science • D2.2 - Ability to address and solve a problem using scientific methods • D2.6 - Ability to apply innovative techniques of data mining and machine learning to extract knowledge from complex and heterogeneous data <p>Making judgments</p> <ul style="list-style-type: none"> • D3.2 - Ability to autonomously select the documentation (in the form of books, web, magazines, etc.) needed to keep up to date in a given sector <p>Communication skills</p> <ul style="list-style-type: none"> • D4.1 - Ability to use English at an advanced level with particular reference to disciplinary terminology • D4.3 - Ability to structure and draft scientific and technical documentation <p>Learning skills</p> <ul style="list-style-type: none"> • D5.2 - Ability to autonomously keep oneself up to date with the developments of the most important areas of data science
--	--

Assessment	<p>The assessment of the course consists of three parts:</p> <ul style="list-style-type: none"> • Midterm Written Exam (optional) • Final Written Exam • Projects: assessed on group-based project assignments <p>Non-attending students should take the midterm and final written exams and work on an independent project to be agreed with the lecturer.</p>
Assessment language	English
Assessment Typology	Monocratic
Evaluation criteria and criteria for awarding marks	<p>The learning outcomes are assessed by means of written examinations directed to identify the first two areas (knowledge and its application), while the group-based project work on a given practical problem is directed to the assessment of the latter areas.</p> <p>The assessment is based on group-based project assignments and written examinations.</p> <p>The final mark will be formed as follows:</p> <ul style="list-style-type: none"> • Final Written Exam: 70% • Projects: 30% <p>Optional: the students may take a midterm exam, which will count as 50% of the written examination, i.e., 35% of the final mark.</p> <p>Non-attending students should contact the lecturer to get to know the date of the midterm exam. They are assessed with the midterm and final written exams (35% for each exam) and the independent project (30%). The final written exam will cover all the course. Those students who did the midterm exam will have the option of only doing the second half of the final exam (covering the content of the second half of the course).</p>

	<p>Project assignments are proposed during the course and delivery procedure and deadline will be announced on the course website and in class. The evaluation of the project is based on the group results and the individual contributions. Students will be asked to prepare a presentation, write a report, and provide the code and executables related to the project.</p>
<p>Required readings</p>	<p>Selected chapters from:</p> <ul style="list-style-type: none"> • Chris Manning and Hinrich Schütze, <i>Foundations of Statistical Natural Language Processing</i>, MIT Press, 2003. (Unibz Library Code: ES 910 M283) • Liu, Bing. <i>Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data</i>. Springer, 2011. • Feldman, Ronen and Sanger, James. <i>Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data</i>. Cambridge University Press, 2006. <p>Subject Librarian: David Gebhardi, David.Gebhardi@unibz.it</p>
<p>Supplementary readings</p>	<p>Additional material will be provided during the course.</p>
<p>Software used</p>	<p>--</p>