## COURSE DESCRIPTION – ACADEMIC YEAR 2018/2019

<table>
<thead>
<tr>
<th>Course title</th>
<th>Web and Text Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course code</td>
<td>73015</td>
</tr>
<tr>
<td>Scientific sector</td>
<td>INF/01</td>
</tr>
<tr>
<td>Degree</td>
<td>Master in Computational Data Science (LM-10)</td>
</tr>
<tr>
<td>Semester</td>
<td>1</td>
</tr>
<tr>
<td>Year</td>
<td>1</td>
</tr>
<tr>
<td>Credits</td>
<td>6</td>
</tr>
<tr>
<td>Modular</td>
<td>No</td>
</tr>
</tbody>
</table>

| 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/](https://ole.unibz.it/)

**Specific educational objectives**

The course belongs to the type "caratterizzanti – discipline informatiche" in the curricula “Data Analytics” and “Data Management”.

In this course the students will develop an understanding of fundamental natural language processing techniques used for acquiring and representing knowledge from the Web.

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.

**Lecturer**

Ana Ozaki

**Contact**

Piazza Domenicani 3, Room 3.04, [ana.ozaki@unibz.it](mailto: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**

- 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:
• 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

Making judgments
• 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

Communication skills
• 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

Learning skills
• D5.2 - Ability to autonomously keep oneself up to date with the developments of the most important areas of data science

Assessment

The assessment of the course consists of three parts:
• Midterm Written Exam
• Final Written Exam
• Projects: assessed on group-based project assignments

Non-attending students should take the midterm and final written exams and work on an independent project to be agreed with the lecturer.

Assessment language
English

Assessment Typology
Monocratic

Evaluation criteria and criteria for awarding marks

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.

The assessment is based on group-based project assignments and written examinations.

The final mark will be formed as follows:
• Midterm Written Exam: 35%
• Final Written Exam: 35%
• Projects: 30%

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.

Non-attending students are assessed with the midterm and final written exams (35% for each exam) and the independent project (30%).
## Required readings

Selected chapters from:


Subject Librarian: David Gebhardi, [David.Gebhardi@unibz.it](mailto:David.Gebhardi@unibz.it)

## Supplementary readings

Additional material will be provided during the course.

## Software used

--