# COURSE DESCRIPTION – ACADEMIC YEAR 2018/2019

<table>
<thead>
<tr>
<th><strong>Course title</strong></th>
<th>Web and Text Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Course code</strong></td>
<td>73015</td>
</tr>
<tr>
<td><strong>Scientific sector</strong></td>
<td>INF/01</td>
</tr>
<tr>
<td><strong>Degree</strong></td>
<td>Master in Computational Data Science (LM-10)</td>
</tr>
<tr>
<td><strong>Semester</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>Year</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Credits</strong></td>
<td>6</td>
</tr>
<tr>
<td><strong>Modular</strong></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/

## 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

## 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:
### Assessment

The assessment of the course consists of three parts:
- Midterm Written Exam (optional)
- 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:
- Final Written Exam: 70%
- Projects: 30%

Optional: the students may take a midterm exam, which will count as 50% of the written examination, i.e., 35% of the final mark.

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).
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.

<table>
<thead>
<tr>
<th>Required readings</th>
<th>Selected chapters from:</th>
</tr>
</thead>
</table>

Subject Librarian: David Gebhardi, David.Gebhardi@unibz.it

<table>
<thead>
<tr>
<th>Supplementary readings</th>
<th>Additional material will be provided during the course.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Software used</th>
<th>--</th>
</tr>
</thead>
</table>