

Information: Classes exams and topics

*Linguistic Resources for Natural Language Processing
LM Language Technologies and Digital Humanities
2024-25*

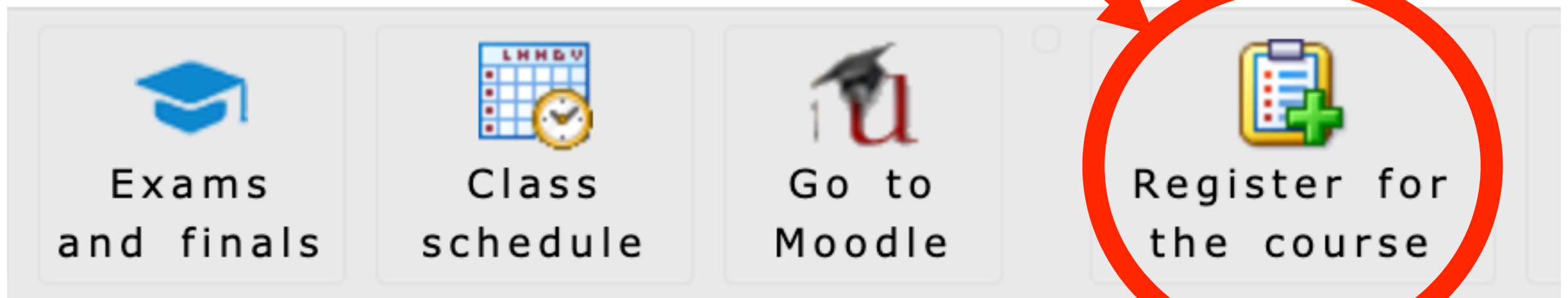
Cristina Bosco

Enrolment and materials

- The webpage of this course is:

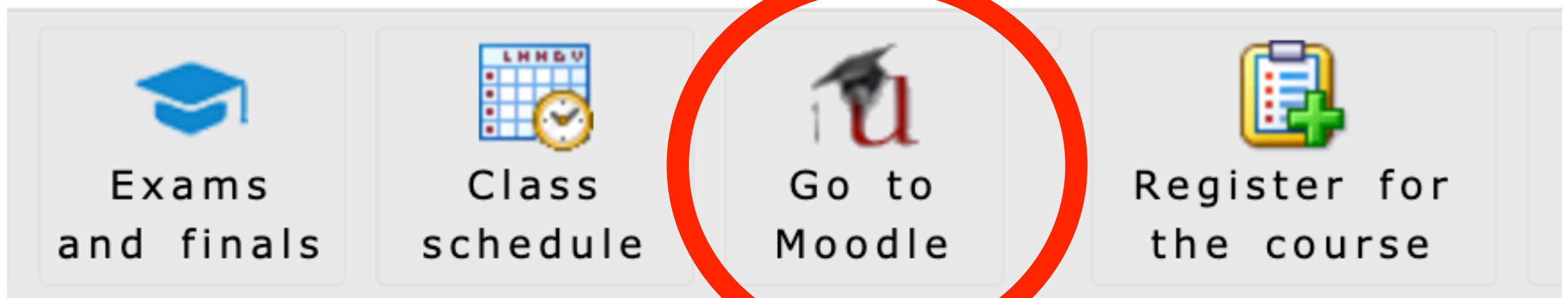
https://www.digitalhumanities.unito.it/do/corsi.pl/Show?_id=nxwv

- Students must **enrol** for this course by logging in with their credentials and selecting “**Register for the course**” at the bottom of the page



Enrolment and materials

- Students can access course **materials** (slides and other files) through the Moodle platform, by selecting “**Go to Moodle**” at the bottom of the page (and then logging in again)
- Students must also **register in Moodle as course participants**



Classes and exams

- This course consists of **two parts** given by two different teachers, **Cristina Bosco (first part)** and **Viviana Patti (second part)**, and some seminars given by experts
- Classes will include frontal **teaching** and **practical activity**
- Practical activities take place also in **groups** and are focused on analysing and building linguistic resources
- The **final exam** will be about the two parts both, and the evaluation will be also based on the participation to the proposed practical activities

Class schedule

FIRST PART (*teacher Cristina BOSCO*)

from September 16th to October 23rd

room 3.06 / from 2pm to 4pm

◆ **Monday**

◆ **Tuesday**

◆ **Wednesday**

Class schedule

SECOND PART *(teacher Viviana PATTI)*

from November 11th to December 18th

room 3.06 / from 2pm to 4pm

◆ **Monday**

◆ **Tuesday**

◆ **Wednesday**

Seminars of FIRST part

◆ **Elisa Di Nuovo (ISPRA): Valico-UD**

October ????th (room 3.06 / from 2pm to 3pm)

◆ **Francesco Mambrini (UniCattolica Milano):
Resources for ancient languages**

October ????th (room 3.06 / from 2pm to 3pm)

Evaluation of the course

The evaluation (maximum 30 points) also is divided in 2 parts:

- 15 points (maximum) for the FIRST part
- 15 points (maximum) for the SECOND part

CLASSES

First PART (Bosco)
Sept 16 - Oct 23

Second PART (Patti)
Nov 11 - Dec 18

EXAM

First PART (Bosco)
In appeal

Second PART (Patti)
In appeal

<15

+

<30

<15

Evaluation of FIRST part

CLASSES

EXAM

First PART (Bosco) Sept 16 - Oct 23	First PART (Bosco) <i>In appeal</i>	<15	= A (max 7)
Second PART (Patti) Nov 11 - Dec 18	Second PART (Patti) <i>In appeal</i>	+ <30	
		<15	

A - score = (maximum) 7 points to evaluate your knowledge about **morphological and syntactic analysis** and ability to use **tools for applying it to texts**

You will be asked to describe in a **talk with slides** the application of these tools to some case-studies for a language you are familiar with, working in group (maximum 4 members) or as a single student

Evaluation of the first part

CLASSES

EXAM

First PART (Bosco) Sept 16 - Oct 23
Second PART (Patti) Nov 11 - Dec 18

First PART (Bosco) <i>In appeal</i>
Second PART (Patti) <i>In appeal</i>

<15 = **A** (max 7) + **B** (max 8)
+ <30
<15

B - score = (maximum) 8 points to evaluate your **knowledge about linguistic resources and their exploitation in different tasks**

You will be asked to **answer an oral question** about a topic selected (by the teacher) in a list that will be distributed before the end of the first part of the course

Goals and map

- To know the **linguistic resources** most commonly used in Natural Language Processing
- To learn how to **build different types of linguistic resources**, starting from scratch or exploiting previously developed sets of data and using tools
- To learn how to **address** the variety of **challenges** involved in the development of a linguistic resource, closely related to different domains and genres, languages, formats

Introducing myself

- I graduated in Philosophy - *Natural language analysis and computer science*
- PhD in Computer Science
- Associate Professor in Computer Science
- I'm a computational linguist and founding member of

the



Associazione Italiana di
Linguistica Computazionale

I coordinated and participated to the development of several **linguistic resources**:

- PoS tagged corpora: PoSTWITA, KiPoS
- Treebanks: Turin University Treebank, Parallel Turin University Treebank, PoSTWITA-UD, VALICO-UD, Twittirò-UD, ISDT, VIT
- Corpora for classification tasks: SentiTUT, HaSpeeDe I, 2 and 3, SemEval2019 task 5, IronIta, Policy corpus XL, FB-Stereotypes
- Lexica: MAL and WMAL

I worked on several different **text genres and domains**

- legal texts, newspapers, Wikipedia, social media, conversation ...
- about politics, climate change, immigrants ...

And for different **languages**

- Italian, English, French, Spanish, Catalan, Indonesian

I worked on several different **shared tasks**, as participant or organiser:

- Dependency and constituency parsing
- Part of speech tagging
- Text classification tasks, e.g. sentiment analysis, hate speech (against immigrants or women) detection, irony detection, stance detection, stereotype detection

My current research projects are mostly about text classification applied in projects where we address societal challenges, such as detecting hate speech and racial stereotypes spreading online or issues related to the environment.



contro ~~l'~~ **podio**



DIREZIONE GENERALE DEL TERZO SETTORE
E DELLA RESPONSABILITÀ SOCIALE DELLE IMPRESE



STERHEOTYPES

Studying European Racial
Hoaxes and Sterheotypes



Finally ... I am a native Italian speaker, I write and teach in English and use **English as the Medium of Instruction** (EMI).

My main goal is to be clear, to look for synonyms to express the content in different ways and to ask questions to understand if the students have understood the content.

Learning is the process by which people acquire knowledge, attitudes and skills.

LEARNING IS A SOCIAL PROCESS

It depends on attention and observation, but above all on **INTERACTION** with the teacher and other students.

QUESTIONS ?????