Co-funded by the
Erasmus+ Programme
of the European Union

UP SKILLS

i. Name of the course

Research-based Course in Multilingual NLP

ii. Level of the course

BA, MA

iii. Workload

6 ECTS

iv. Institution

University of Zurich

v. Course instructor(s)

Tanja Samardžić

vi. Brief course description

The students enrolled in this course will pursue small individual research projects, each tackling a separate problem as one part of a bigger project. They will be shortly introduced to the main opportunities and challenges of multilingual NLP and then select the problem to work on during the course. For example, the question of what tokenization method is optimal for which script requires studying several scripts and several tokenization methods. Each student can work on one script comparing several tokenization methods.

At the end of the course, each student will have written a 10-page research report following the same template. During the course, the students will regularly submit in-progress reports and meet with the lecturer in weekly interactive sessions.

The main output of the course activities will be students' final reports. In addition to this, the lecturer will compile a progress report summarising observations on how the course advanced each week. The lecturer's report will contain an overview of the students' projects and how they relate to each other.

vii. Research related subject

Multilingual NLP

viii. Tools and data the students work with

Own Python programs, data from highly multilingual data sets (our own TeDDi sample, other data sets derived from Wikipedia)

ix. Topics

See “Learning outcomes”.

x. Learning outcomes

The learning outcomes of this course are divided into two parts, topic-specific and transversal. Regarding the topic-specific part, the students who complete this course will have an overview of the issues related to multilingual processing such as the impact of various writing systems on text encoding, subword tokenization and basic text statistics. They will be able to apply relevant steps in the current processing pipeline (using large pre-trained models and cross-lingual transfer). Regarding the transversal skills, the students will gain experience with problem solving, forward thinking and independent work. They will activate their own curiosity as a driver of research activities.

xi. Evaluation

Research report evaluation, following the criteria outlined in the UPSKILLS guidelines.