

NLP for low-resource languages

Syllabus

WS 23/24

Instructor: Mareike Hartmann (mareikeh@lst.uni-saarland.de)

Meetings: Wednesdays 10:15 - 11:45, C7.3 Room 1.12

The term 'low-resource language' is an umbrella term used to refer to a large amount of languages that have very different amounts and types of resources available, ranging from languages with millions of speakers but no curated datasets to train supervised NLP systems, to languages with a handful of speakers and no written records at all. In the first part of the seminar we will talk about different ideas of how to define low-resource status and what causes and implication low-resource status has. Then we will talk about NLP initiatives for specific language communities, e.g. resource collection and general recommendations on successful workflow. And then we will read some papers that are representative of the main approaches that NLP researchers take for processing low-resource languages.

Prerequisites: Background in natural language processing and deep learning is required.

Course Requirements:

- **Read papers, hand in questions, and actively participate in the discussion.** You are expected to read the papers to be presented in class beforehand, and actively participate in the discussion during class. You are asked to send 1-2 questions and/or comments related to the assigned paper by 19:00 the day before class. I will share these questions with the presenter so they can be integrated into the discussion.
- **In-class presentation.** Each student has to give a presentation about one of the listed papers and lead a discussion based on the presented paper. The presentation should last 30-40 minutes and be followed by a discussion for the remainder of the class. You are welcome to integrate other activities (e.g. presenting a demo or providing a tutorial) into your discussion. In preparation of your presentation, you are asked to go over the slides with me at least one week prior to the session. You will be assessed on comprehension of the material and preparation to lead critical discussion of it.
- **Term paper (only for 7CP option).** You have two options for the structure of your term paper:
 1. Expand on one of the topics discussed in class. Note that it is **not** enough to provide a written summary of a paper presented in class. You could e.g. do a comparison of methods or other aspects across papers, provide a survey of a

- specific topic, raise a question and discuss/answer it based on the literature.
2. Do a small project related to the topics discussed in the seminar. For this option, getting in touch with me to discuss the topic and scope of the project is mandatory.

Format of the paper/report: 8 single-spaced pages following the ACL 2022 template

Hand-in date: end of semester (April 7, 2024)

Grading:

- 7CP option: 15% class participation, 30% presentation, 55% term paper
- 4CP option: 34% class participation, 66% presentation