

Subgroup Report: Documentation Task (A. Seza Doğruöz)

- During the Istanbul meeting (September 2023), we discussed the available platforms for language resources (both data sources & tools).
- After evaluating the available ones, we have decided that the European Language Grid (ELG) is the one that will serve our purposes (However, one participant has faced issues with it and I am not sure what is going on).
- There are gaps in the meta-data and WG3 can collaborate with ELG toward better documentation.

Survey

As sub-group leaders, we have piloted a survey among the participants of the UniDive Istanbul meeting to understand their academic backgrounds (e.g. linguistics, computational linguistics, others), knowledge about the language resources, platforms, tools, which languages they work on, expectations and motivations.

This pilot survey was well-received within the Istanbul group. In addition, we have received requests beyond WG3 (by other WGs) to extend this survey to all UniDive members. We are currently working on the improved version and we will keep you updated about it soon. Maria (ELG) is also joining us for further help with ELG.

What to do next?

- Initial Idea: Collaborate on scientific articles (e.g., surveys) about multilingual tools & resources (also covering ELG).
- We can use our survey can help us to collect data on different tools & resources.
- In addition, we can check them through search engines and find out ways how to complement this information with our survey (suggestion by Kristina).

However:

- We have not resolved the issue about what type of multilingualism we are talking about? I have raised this issue in the Istanbul meeting, one of the participants remembered that and commented on it (good point). In addition, there are multiple tasks (across WGs) working on this issue. We are doing similar things. Perhaps, we can unify these efforts instead of creating new tasks.
- There is a special issue coming up on LRE about Language Technology Platforms (by ELG people) which will probably cover multilingual tools & resources.
- There are also in-depth studies by ELG covering the language resources & tools for European Languages (albeit mostly focusing on 1 language from a monolingual perspective)
- How about the LLMs? Where do we place them while listing the existing multilingual tools & resources?