WG3 1st Meeting Minutes - 2023-03-17

Session 1

- 10.45–11.00 Introduction to WG3 (slides)

- 11.00–11.30 Brainstorming on ideas and expectations

Discussion questions:

- What is most important for you in multilingual and cross-lingual NLP?
- What activities do you think we should prioritize?
- How can we work together to make progress towards our goals?

Points raised:

- Large language models are most important
- Articulating linguistic theories underlying tools
- Defining idiosyncrasy and diversity
- The user perspective is important
- Supporting low-resource languages through cross-lingual technology
- Supporting low-resource languages through annotation tools
- Supporting low-resource languages through data collection
- Supporting low-resource languages with semantics
- Tools for all languages – start with morphology
- Low-resource language is not a homogeneous concept
- Building resources for specific languages (Serbian)
- Linking corpus resources between languages
- Standardized tools applicable to different languages
- Evaluation of tools – coordinate with other WGs
- Tracking evaluation status for different types of tools
- Improved benchmarking and experimental design
- Organize shared tasks

```
* 11.30—12.00   Initial discussion on documentation of tools
```

Discussion questions:

- Which types of tools do we want to include?
- Where do we want to keep the documentation?
- How do we create this documentation/inventory?

Points raised:

- A huge multidimensional matrix

- A shared repository
- Tools shared between typologically similar languages
- Consider end users
- Too many languages have nothing – document what is missing rather than what exists
- Connect to CLARIN
- Flagship project on MWE
- Include all tools or be selective?
- What about commercial tools?
- What about tools without documentation?

WG tasks emerging from the discussion:

- Define multidimensional taxonomy of tools for documentation
- Define infrastructure and procedure for creating documentation