Scavenger Hunt - Let's learn RegEx the fun way!

Barbara Kovačić, Computational Linguistics at Ludwig-Maximilian-University Munich

When working on linguistic databases such as the Corpus Query Processor (CQPweb)[1] or grewMatch[2], it is easy to find specific words or phrases by entering them into the search field or specifying search parameters in the advanced search options. But as soon as we try to search words which fulfil a specific requirement or follow a specific pattern, querying the databases can become time consuming, e.g. searching words which end with the suffix "ing". Fortunately, most linguistic databases offer the possibility to search after so-called "Regular Expressions". Regular Expressions (also known as "RegEx") are a versatile tool for pattern matching and text manipulation. They enable the user to define specific patterns or sequences of characters, making it easier to search for, extract, or replace text within a document or dataset.

Therefore, this workshop aims to demystify RegEx through gamification. By doing so, the participants of the workshop will first get a short introduction in the main elements of RegEx, followed by a scavenger hunt, where they will use their acquired RegEx knowledge to solve a linguistic riddle.

Along the way, they will learn:
- Syntax and Building Blocks
- Patterns and Wildcards
- Quantifiers and Groups
- Alternation and Logical OR
- Anchors and Boundaries
- Lookaheads and Lookbehinds

References

[1] *CQPWeb main page*. (n.d.). CQPweb. https://cqpweb.lancs.ac.uk/
[2] *Grew-match main page*. (n.d.). https://match.grew.fr/