**Project title:**

Enhancing and improving accuracy of the Ukrainian UD treebank using resources of the Ukrainian Brown Corpus.

**Name and affiliation:**

Kyrylo Popik, Ukraine, National Technical University «Kharkiv Polytechnic Institute», Department of Intelligent Computer Systems, MS student.

**Keywords:**

Treebank, Accuracy, Ukrainian, Ukrainian brown corpus.

**Resource description:**

The current Universal Dependencies treebank for Ukrainian contains 122k tokens in 7000 sentences of various genres, with ~91% annotation accuracy. To improve its accuracy, adding more tagged tokens to the corpus is required.

There exists an annotated corpus of Ukrainian language by the name of Ukrainian Brown Corpus (https://github.com/brown-uk/corpus), which can be used to enrich the UD treebank with new, quality tokens. However, the tagging used there is based on V. Large Electronic Dictionary of Ukrainian (VESUM), and such, there is a need for tag conversion, as those do not match the UD token tags 1 to 1.

The end result of the project is intended to be an annotated corpus of Ukrainian tokens currently absent from Ukrainian treebank, in UD tagging style, which can be later merged to improve the accuracy of Universal Dependencies treebank for Ukrainian language.

**How participation will be useful:**

Construction of an improved treebank requires additional knowledge in fields of syntactic annotation, UD and SUD, annotation platforms for smooth and speedy progress.

The courses featured in this training school deal with the subject of Universal Dependencies corpus annotation and the infrastructure used for annotation, knowledge that is crucial to building an improved treebank and thus invaluable to completion of this resource construction project.

## Questions:

How can one create an efficient framework for converting a corpus into a different annotation scheme that reduces the need to manually audit each converted token?

What could be a good way to automate the testing of new treebank data for the purposes of refining a new resource?

## Project phase:

Project started, in research phase, testing various software approaches.