Automated test generator for Estonian language

The aim of this project is to develop a tool, which will automatically generate tests out of existing text corpora, which simplify training and assessment of Estonian language skills. The tests are constructed by applying a natural language processing tool to identify parts of speech, which will be left blank in tests. The tests are packaged in H5P format and are deployed in the Web.

Background

Substantial amount of time in language education is spent on preparing tests for training and assessment of students’ language skills. Often this is done by taking a text from an online source, relevant parts of speech to be tested are identified and replaced with blanks. Finally tests are published in online environments, where students need to fill in the blanks.

In order to significantly improve language teaching, there should be tons of such tests available, which vary in length, level of difficulty, parts of speech to be trained/tested and vocabulary. Furthermore, in order to reduce cheating, especially in distance learning settings, there should be even more tests available at any time to keep test reuse rates low and thereby incentive to share test answers among students low as well. Currently only few of such tests are available.

Impact and beneficiaries

As a result of this project availability of publicly available online learning material for Estonian language will increase significantly. The main beneficiaries of the project are Estonian language instructors (get a tool, which simplifies test generation) and students (availability of tests, which match the needs and preferences of specific students).

MVP scope

- generator is executable from command line
- generator takes as input plain text file and parameters identifying which parts of speech (noun, pronoun, verb) and in which form should be “left blank” for testing
- generator applies Estonian natural language processing toolkit to identify in the text parts of speech, which match the parameters set
- generator generates learning material of H5P.Blanks content type with test assessment criteria and automated feedback
- generator packages tests in H5P format.

Extra credits are given for efforts, which enable: 1) Web-based front-end for test generation, 2) automated filtering of text files (for bulk test generation) from a set of text files by text length (in words), frequency of matching parts of speech elements and difficulty of words (wrt given database of words), 3) automated deployment of generated tests in an xAPI-enabled Drupal environment, 4) automated publishing of learning material metadata at E-koolikott.

See also

- Tools for Estonian language teachers (see https://sonaveeb.ee/teacher-tools/#/grammar shows which data is available for mappings from forms of nouns, verbs and pronouns to learning outcomes plus defines language skill levels wrt learning outcomes and vocabularies.)
- Estnltk – Open source tools for Estonian natural language processing (see https://estnltk.github.io/estnltk/1.4.1/tutorials/morphology_tables.html for a listings of forms of nouns, verbs and pronouns.)
- Personalized learning path infrastructure project

Contact

Peep Küngas (peep.kungas@harno.ee)