

Ptakopět v2

Překlad tam a kontrolně zpět

Specification

Vilém Zouhar, Ondřej Bojar (advisor)

Feb 2019
Rev. 2

1 Overview

Ptakopět v2 aims to extend the functionality of Ptakopět v1 as part of the Semestral Project in the summer semester of 2018/2019. Ptakopět is a sample user interface for the so-called Outbound Translation, allowing the user to produce text in a language they don't speak. The goal of the extension is to provide further information to the user in order to improve the confidence of the user that the translation to the target and unknown language is correct. The main focus will be on utilizing a quality estimation model.

2 Phases

The development goals are split into three phases:

2.1 Robust interface

Ptakopět v1 was developed as a browser agnostic plugin. This resulted in several technical issues which we were not able to resolve (most notably custom input elements). The basic functionality will be migrated to a static page ptakopet.vilda.net. This will resolve the problem of system's dependency on third party websites.

The general applicability of Ptakopět gets somewhat decreased by this change (the user can no longer use Ptakopět on any web page) but the implementation of Ptakopět will remain as independent of the web page as possible. In other words, adding Ptakopět functionality to a web page containing a text form will be still very easy for *the web page owner* by a small modification of the web page source but no longer for the web page user.

2.2 Quality Estimation Model

We will find and set up a quality estimation model with exposed API to be used in Ptakopět. We will utilize some of the models which participated in [WMT Quality Estimation Task](#) (eg. [Quest++](#)).

2.3 Quality Estimation User Interface

We will explore possibilities of passing relevant quality estimation information to the user in Ptakopět v2. For this part we may need to utilize some sort of word alignment. Relevant systems are: [fast_align](#), [MGiza](#) and [Giza++](#).

The first UI implementation should include color underlining based on the quality estimation model output values. Other alternatives will be highlighting words with quality value below some threshold. Both approaches should provide a fluent and non-obstructing translation quality estimation service to the user.