Paraphrase Generation for Czech

Veronika Burgerová

April 15, 2020



2 The Watsonson Project

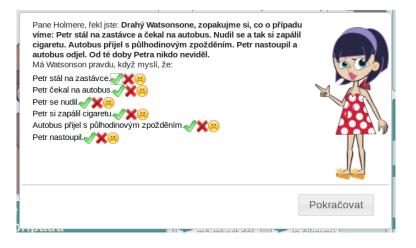


- Paraphrase generation useful for several NLP applications
 - Machine translation
 - Text summarization
 - ...
- Rule-based and machine learning methods
- Improvement of rule-based paraphrasing system for Czech

- By Zuzana Nevěřilová
- A game with purpose
- Generates paraphrases and entailments from a given sentence

The Watsonson Project

Shenlock Holmer a Dr. Watsonson pátrají společně Hra Drahý Holmere, jaký případ právě řešíte? Úroveň: 0 Body: 0 Detektiv: Jméno Watsonsonův mozek Paměť Hypotézy Řekněte Watsonsonovi o novém případu Řekněte dr. Watsonsonovi, že Dedukční schopnosti 1 má mluvit dál je šikovný 🚩 je nemožný × mu poradíte 🛛 🚩 se vrátíte ke starému případu 🛛 **E** mu popíšete nový případ už nemáte čas ani trpělivost



- synonym replacement
- hypernym replacement
- word reordering
- anaphora resolution
- verb frame inference

- Incorrect sense in word replacement
- Errors caused by syntactic parser

• New module in progress - Voice transformer

- Transforming active voice to passive
- Working for simple sentences so far
- Generates only one phrase (not always the correct one)
- Example:
 - Input: "Babička upekla sladkou buchtu. Ten starý muž nehledá ženu."
 - Output: "Babičkou byla upečena sladká buchta. Tím starým mužem není hledána žena."

- Reimplementation of the module for synonym replacement
- Implemented in Python 3
- Generates all possible combinations
- Example
 - Input: Tohle už bohužel bude složitější.
 - Some of the outputs:
 - Tohle už naneštěstí bude složitější
 - Tohle už bohužel bude komplikovanější.
 - Tohle už naneštěstí bude spletitější.
 - ...

- Doing some research about this topic
- Trying bidirectional language model
- Approaches being used:
 - Computing probability of the replaced words in given context
 - Computing probability of the whole sentence
- No results yet

- Use the model to get rid of no sense phrases
- Add syntactic parsing to synonym replacer
 - To be able to replace with not grammatically similar words
- Improve the voice transformer
 - Generate all the possible phrases
 - Fix some errors
 - Make it work for more complex sentences
 - Eventually add the passive-to-active part
 - Add more features (reflexive passive, transform only transitive verbs...)