

Natural Language Processing by Utilizing Crowds

Jozef Harinek
Marián Šimko

j.harinek@gmail.com

25.11.2014

What is it about?

NLP (Syntactic analysis - Parsing)

Crowdsourcing

Problem

Obtaining annotations.

It is hard to get automated annotations for Slovak language.

- irregularities, not much research

Crowdsourcing, we believe, will help.

[1] Čižmár, A., Juhár, J., Ondáš, S. (2010): Extracting sentence elements for the natural language understanding based on slovak national corpus. In Proceedings of the International Conference on Analysis of Verbal and Nonverbal Communication and Enactment, COST'10, LNCS Vol. 6800, Springer, 2010 pp. 171-177.

[2] Marta Sabou, Kalina Bontcheva, and Arno Scharl. 2012. Crowdsourcing research opportunities: lessons from natural language processing. In Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW '12). ACM, New York, NY, USA, , Article 17 , 8 pages. DOI=10.1145/2362456.2362479 <http://doi.acm.org/10.1145/2362456.2362479>

Hypothesis

Utilizing crowdsourcing methods for obtaining annotations will:

- have better results than automated methods.
- give same quality annotations, but with lesser effort than manually created by experts.

Crowdsourcing dimensions

Motivation

Quality control

Aggregation

Human skill

Process order

Task-request cardinality

[2] QUINN, Alexander J.; BEDERSON, Benjamin B. Human computation: a survey and taxonomy of a growing field. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011. p. 1403-1412.

Crowdsourcing dimensions

Motivation - points for final grade

Quality control - output agreement, teacher

Aggregation - collection

Human skill - syntax analysis knowledge

Process order - Requester-Worker

Task-request cardinality - many-to-many

Data

Dataset created by teachers when creating assignments.

Dataset from Slovak Academy of Science.

Idea validation

Two groups of 20 students.

One session of sentences annotation with each.

Evaluation in next term

Extensive experiment.

Several schools - many classes.

Teacher-friendly environment (task specification, automatic correction).

Expectations & Conclusions

Proving the concept as it is set up in crowdsourcing.

We will obtain valid annotations.

We will obtain same quality annotations more easily by utilizing the power of crowd instead of experts.