University *of Ljubljana*
Faculty *of Computer and Information Science*

Literacy situation models knowledge base creation

Nace Gorenc, Jernej Vrhunc in Žan Pečovnik

Abstract

Advisors: Slavko Žitnik, Aleš Žagar

Introduction

In recent years natural language processing became one of the hottest research topic. One of the driving forces was public availability of ChatGPT, which caused quite a stir since people started using it for various task which were formerly done by hand and with a little bit of help from search engines like Google. This report focuses on the analysis of personal relationships between fictional characters. (js sm tako zacel, lahk se se kej doda do srede ko bomo mel se kej spisano)

Related work

A relation extraction boggled minds of the scientists long before NLP's renaissance driven by the invention of transformers.

In 2011, authors of [1] reviewed multiple supervised and semi-supervised classification approaches to the relation extraction task along with critical analysis. Among the supervised approaches, dependency path kernels stood out as the best both in term of computational complexity and performance. On the other hand, semi-supervised approaches seemed to be well suited for open domain relation extraction systems since they could easily scale with the database size and could extend to new relations easily.

Groza and Corde [2] focused on extracting information about literary characters from folktales. One of the tasks was to identify the main characters and the parts of the story where these characters are described or act. Their system relies on the folktale ontology which is based on Propp's model for folktales morphology and it defines three knowledge sources. The first one is *folktale morphology*, the second one is *folktale main entities* and the last one is *family relationships in folktale*. The authors used three sequential algorithms for retrieving information about characters. The main goals of these algortihms were to extract characters from the folktales, use decoreference resolution and finding the perspective of the characters.

More recently the authors of [3] proposed method of a hybrid approach which combines the features of unsupervised and supervised learning methods, which also uses some rules to extract relationships. The method identifies the main characters and collects the sentences related to them. Then these sentences are analyzed and classified to extract relationships. For testing purposes, 100 short stories were used with around 300 character pair relations. This model identified parent-child relationships, friendships or if there was no relation.

Dataset analysis

Methods

Data pre-processing

Results

Future directions and ideas

Discussion

References

[1] Nguyen Bach and Sameer Badaskar. A review of relation extraction. 05 2011.

[2] Adrian Groza and Lidia Corde. Information retrieval in falktales using natural language processing. In *2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 59–66. IEEE, 2015.

[3] V. Devisree and P.C. Reghu Raj. A hybrid approach to relationship extraction from stories. *Procedia Technology*, 24:1499–1506, 2016. International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015).