Entity Network Construction from English novels

Bingjie Jiang
bingjie3@illinois.edu

Ji Li
jili3@illinois.edu

Liyuan Liu
ll2@illinois.edu

Background

Character map, which records interconnected relationships among characters, has been identified as a very helpful tool for not only readers, but also writers. Lots of instructions are available to build character maps through tedious human effort [1], and massive character maps have been created for some classical novels in these ways, e.g. The Tempest [2].

On the other hand, with recent advances in the field of information extraction, many toolkits like Stanford Named Entity Recognizer [1] have been released for recognizing entities. Besides, many efforts have been made to identify and categorize the relations among these entities [3]. Being Aware of the effectiveness of these techniques, we aim to make automatical construction of character maps from text possible. Furthermore, the maps we'd like to construct can include characters and more entities, geographic locations and organizations for instance.

Similar tasks have been conducted on web pages to extract information from them[2]. However, to the best of our knowledge, there is no existing tools designed for our goal, i.e., identifying and extracting entities and their relations from a specific novel. The key difference between our task and other tasks is that, we focus on identifying these relations for a specific text with limited size, while other tasks tend to extract information from web pages with nearly unlimited size. And more meticulous approaches and visualization strategies can be applied on our project.

Details & Plan

Here, we propose a concept mapping construction system, to identify characters, geographic locations and organizations, and the relations among them for a specific text data. We believe it would help readers to gain a clearer view of the books they want to read.

As for the technique part, we plan to leverage the Stanford NER [1] tool to detect entities from text, and employ distributed representation techniques to identify the relation among them, which could capture the semantic meaning and can generalize to other texts.

For demonstration, we would show the process and results of our system. We plan to complete the entities identification part by the mid of March, and then use the rest of time to complete relation extraction, visualization, etc. And Bingjie Jiang is our groups' coordinator.

References

[1] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd annual meeting on association for computational linguistics*, pages 363–370. Association for Computational Linguistics, 2005.

[2] Heng Ji and Ralph Grishman. Knowledge base population: Successful approaches and challenges. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1148–1158. Association for Computational Linguistics, 2011.

[3] Xiang Ren, Zeqiu Wu, Wenqi He, Meng Qu, Clare R Voss, Heng Ji, Tarek F Abdelzaher, and Jiawei Han. Cotype: Joint extraction of typed entities and relations with knowledge bases. *arXiv preprint arXiv:1610.08763*, 2016.

[1] http://writeonsisters.com/writing-craft/6-easy-steps-to-great-character-mapping/
[2] https://www.cliffsnotes.com/literature/t/the-tempest/character-map