Fine-Tuned Relationship Extraction for Consumer Goods Concepts

Data Science Institute
COLUMBIA UNIVERSITY

Wenxin Zhang, Xingyu Lu, Yinghao Li, Jiazhen Li, Wei Luo,
Mentor: John Labarga

Unilever

Data Science Capstone Project with Unilever

Background:

The foundation of Machine Learning models is data, and annotating text data for training NLP models is always time consuming. Existing solutions are usually too complex to use. For example, the annotation tool shown in Fig. 1 has lots of buttons, and users need to manually input the sentences, switch back and forth between several tabs to make a single annotation.

Our project aims to create an open-source annotation software which identifies the entities, recognizes the entity pairs and their corresponding relations using previous models to boost the annotation process. In consequence, annotators could focus on uncertainty of current predictions ,reducing cost both timely and financially.

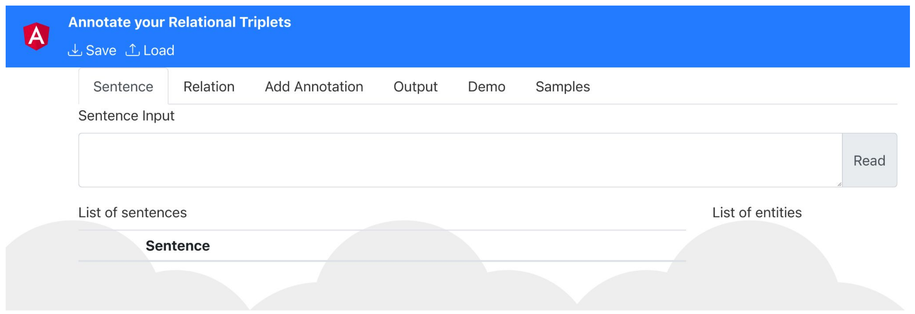


Figure 1. Existing Annotation Tool(https://abera87.github.io/annotate/)

Methods:

Separating frontend and backend is conductive to developing user-interactive functions and making backend focusing on dealing with data when we use models from Open NRE and spaCy. We choose flask, a lightweight framework which supports single page applications and provides full control for users, as our backend building tool. For frontend framework, we choose React due to a steep learning curve. From the perspective of performance, React uses virtual DOM and only loads the different parts in the DOM tree after interacting with users, leading a fast loading speed.
Fig. 2 demonstrates our wireflow design for the front end.

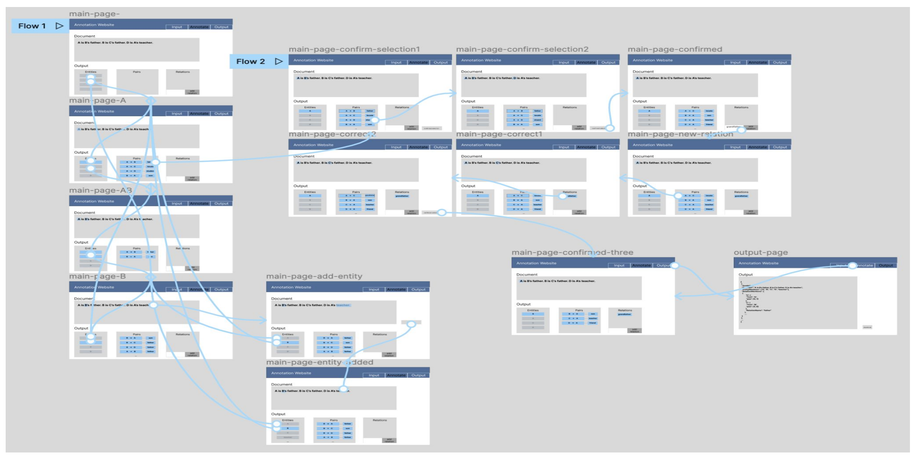
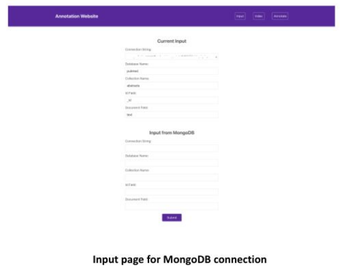


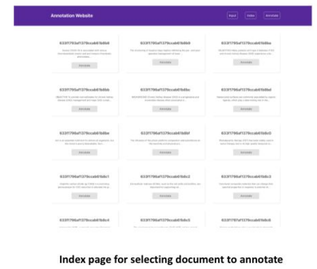
Figure 2. Our wireflow design for frontend

Results

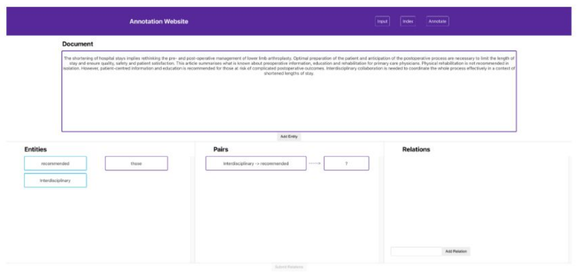
As is shown in Fig. 3, our tool has three main pages: 1) input page that enables users to input documents through mongoDB connection 2) index page that allows users to select documents to annotate and 3) annotation page that unifies all functions like adding entities, relations, assigning relations and etc.



Input page for MongoDB connection



Index page for selecting document to annotate



Annotation page that unifies all functions

Figure 3. Main functions of the annotation tool

Conclusion

Up to now, we have finished implementing the tool and received positive feedback from our monitor. However, whether we need to further improve our website will be depended on the feedbacks from future users. We hope our annotation tool can help boost the process of annotating text contents and make positive impact for the development of NLP field.

Acknowledgments

This project is supported by Unilever. We would like to express our sincere thank to all those who are related to this project. Primarily, we are grateful to our mentor John Labarga and professor Sining Chen for their support and guidance in completing this project.

References

1. https://reactjs.org/
2. https://vuejs.org/
3. https://angular.io/
4. Flask vs Django: Which python framework to choose? InterviewBit. (2022, July 29). Retrieved November 29, 2022, from https://www.interviewbit.com/blog/flask-vs-django/#:~:text=Django%20is%20a%20full%2Dstack,Working%20of%20Framework%2FData%20Model
5. Martin, S. (2022, February 3). Angular vs react vs Vue.js: Which is the best choice for 2022. is. Retrieved November 29, 2022, from https://plainenglish.io/blog/angular-vs-react-vs-vue-js-which-is-the-best-choice-for-2022