

A Visualized Toolkit for Crowdsourcing NLP Annotations

Congle Zhang, Haichen Shen, Hanchuan Li, Shengliang Xu

Research Question: How to use visualization to improve the clustering annotation and tree building tasks efficiency in natural language processing annotation?

Motivation: Manual annotation for NLP training data is well-known for its tedium. In this project we propose a visualization tool improving the efficiency in manual clustering annotation and tree building labeling process for crowdsourcing NLP annotation.

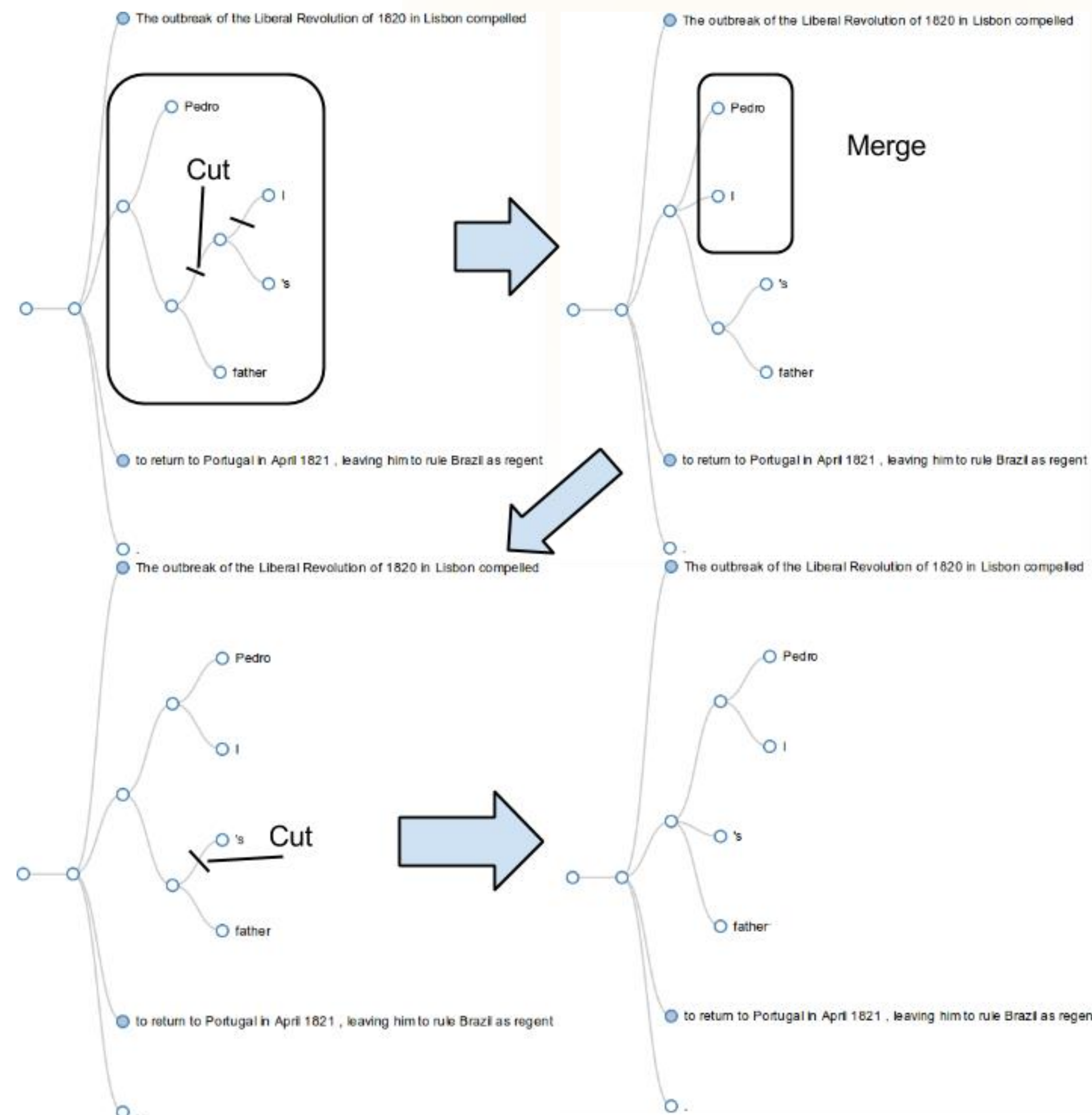
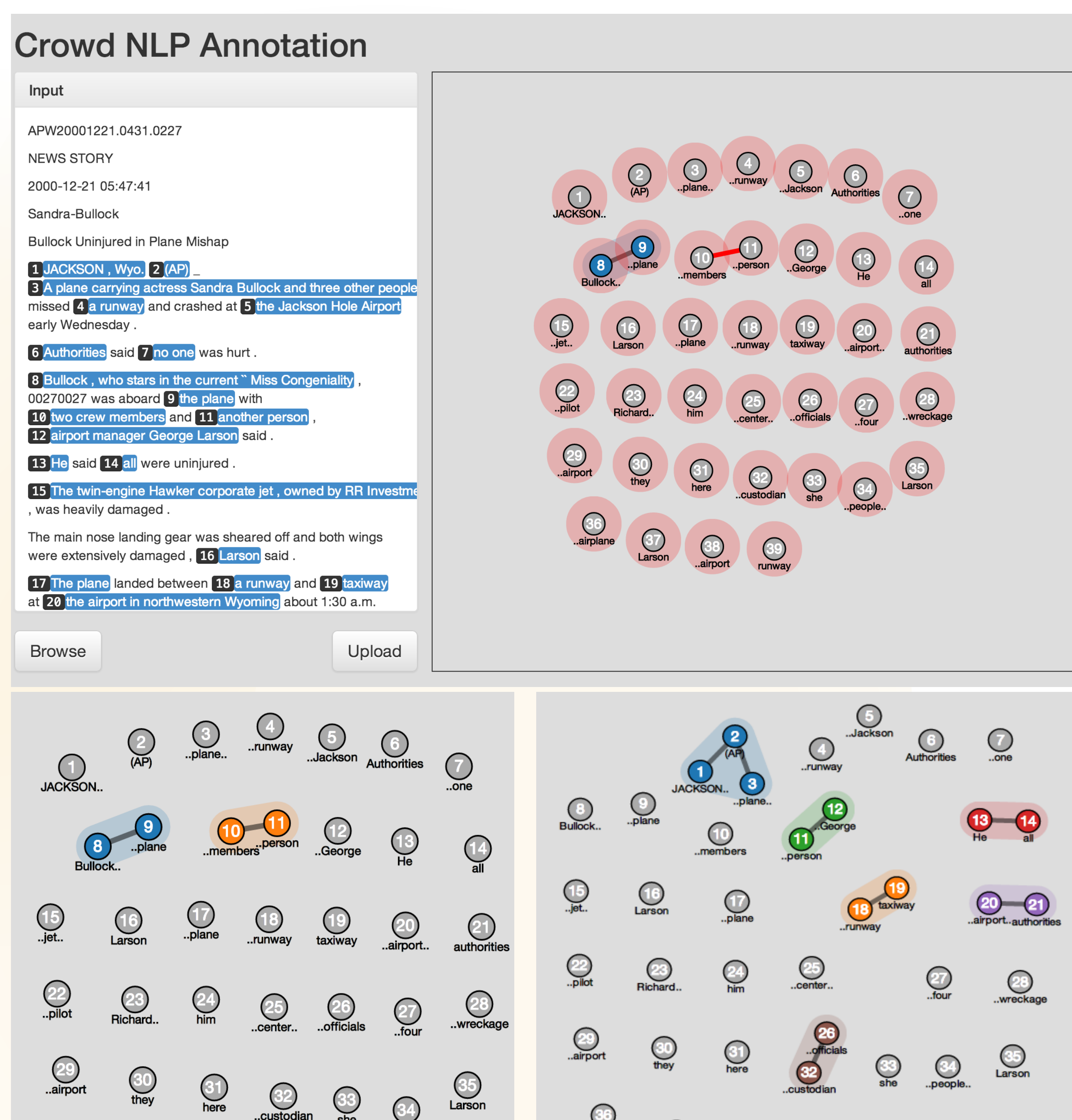
Approach: A web based interactive visualization tool for collaborative annotation focused on following 3 functions:

[1]. The users can use tow and group function to directly operate on the clustering graph to finish the clustering annotating processing.

[2] a interactive visualization of the in building trees. The users can directly operate on the visualized tree to complete the whole annotating process.

[3] Tree involving feature: building new trees from an existing tree to allow cross-lingual tasks.

System Description:



Evaluation, Conclusion & Future Work

We conducted a pilot user study on the clustering and tree building tasks from 4 participants.

On the clustering task, participants showed a average **32%** decrease in time consumption on the clustering annotation task and a similar annotation accuracy.

On the tree building task, participants showed a average **40%** decrease in time consumption with a improved annotation accuracy. Due to high variation between participants, further study will be conducted.

We can clearly see that our visualization is useful for natural language processing annotation. We will have more participants to evaluate the improvement on clustering and tree building accuracy compared to the gold standard tree building answer. One concern we have is in the tree building task, the result is still far from accurate, less than 70 percentage of the elements are correctly labelled. Future study may involve study on machine learning algorithm to pre define tree structure to better improve the tree building task accuracy.