

A Visualized Toolkit for Crowdsourcing NLP Annotations

Congle Zhang, Haichen Shen, Hanchuan Li, Shengliang Xu

Research Question: How to use visualization to improve the entity clustering and tree parsing accuracy and 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 toolkit to improve the efficiency in manual clustering annotation and tree parsing process for crowdsourcing NLP annotation.

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

[1]. *Clustering Tool: Direct dragging function on the cluster graph to enable fast and direct clustering annotating process.*

[2]. *Tree Parsing Tool: Direct cutting and merging function on the tree graph to enable fast and accurate tree building process.*

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

System Description:

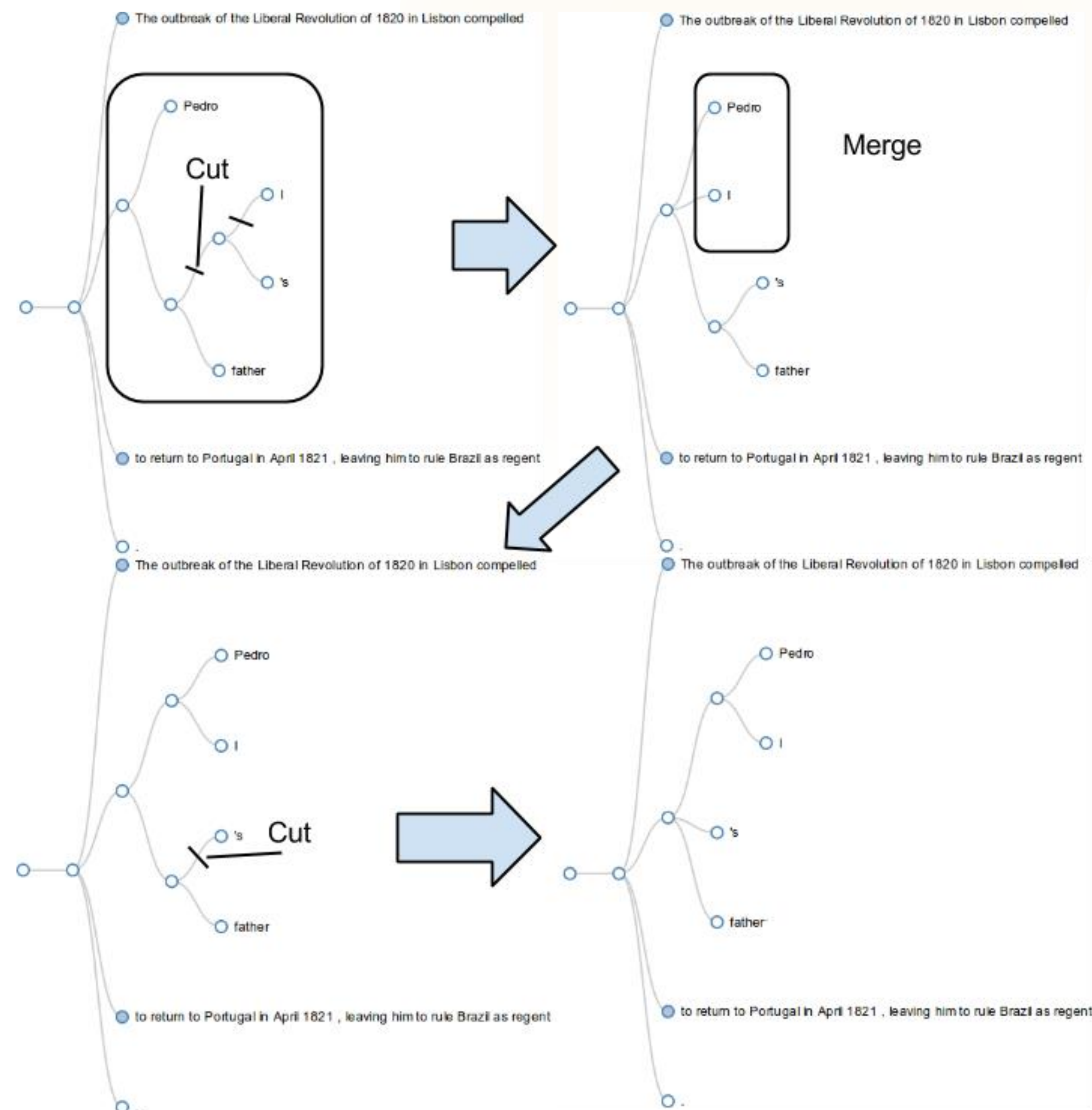
Crowd NLP Annotation

Input

APW20001221.0431.0227
NEWS STORY
2000-12-21 05:47:41
Sandra-Bullock
Bullock Uninjured in Plane Mishap

1 JACKSON., Wyo., 2 (AP) ..plane.. ..runway.. ..Jackson Authorities 7 ..one
3 A plane carrying actress Sandra Bullock and three other people missed 4 a runway and crashed at 5 the Jackson Hole Airport early Wednesday .
6 Authorities said 7 no one was hurt .
8 Bullock , who stars in the current " Miss Congeniality " , 00270027 was aboard 9 the plane with
10 two crew members and 11 another person ,
12 airport manager George Larson said .
13 He said 14 all were uninjured .
15 The twin-engine Hawker corporate jet , owned by RR Investm , was heavily damaged .
The main nose landing gear was sheared off and both wings were extensively damaged , 16 Larson said .
17 The plane landed between 18 a runway and 19 taxiway at 20 the airport in northwestern Wyoming about 1:30 a.m.

Browse Upload



Evaluation, Conclusion & Future Work

We conducted user study on the clustering and tree building tasks from 6 participants.

On the clustering task, participants showed an average **15.2%** decrease in time consumption on the clustering annotation task and with a **6.6%** increase on cluster purity.

On the tree building task, participants showed an average **27.8%** decrease in time consumption with a similar annotation accuracy.

We can clearly see that our visualization is useful for natural language processing annotation. One concern we have is in the tree building task, the result is still far from accurate. Future study may include study on machine learning algorithm to pre define tree structure to better improve the tree building task accuracy and also on quantitative evaluation of the tree building quality.