

Place Descriptions

Arbaz Khan

Progress Report
June 22, 2013

1 Progress

Insight into the Parser Program

The parser program provided by *Felix* was studied to understand its output response to the campus descriptions. The parser program took input in the form of two files containing the POS tagged & chunked text file and the manual annotations in a format proposed by *Igor*. Chunking was done with the help of *Apache-OpenNLP* and the manual annotations were done with the help of *brat* as suggested by *Igor*. The parser program was then run on a short sample input generated from the campus descriptions (I).

This helped getting familiar directly with the complexity of the problem of generating triplets from the output of *Felix's* parser.

Analysis of the functioning

The output provided by the parser is in the form of tagging each word with the IOB encoding. The accuracy of the results on *tellUsWhere* data are observed to be promising. As per the experiments of *Felix*, about three quarters of the locative expressions are recognized by the parser.

However, the *tellUsWhere* data contains pretty straightforward place descriptions where identifying the locative expressions were not non-trivial, esp. as compared to the campus descriptions which contains implicit locative expressions represented by paths. Also, the concern is the input format taken by parser. It requires manual annotations where geospatial noun entities (NE) and noun phrases (NP) need to be identified and tagged for the words occurring in the original place descriptions. Feasibility of the application to a large project stays a major issue henceforth.

2 Future Work

Having set up the input requirements, the next step is to get the output from the parser and visualize the approach for generation of triplets either using DFS method proposed earlier or some other alternative efficient algorithm to extract triplets from the IOB encoded output.