

Information Retrieval Project Report 1

Resolving semantic ambiguities using search engines

Team Members: Brandon Olson, Qiaozhi Li, Bhavana Gudi, Kerwin Zhou, Eric Liu

Github: <https://github.com/kanbei7/IR-Project>

Problem Description

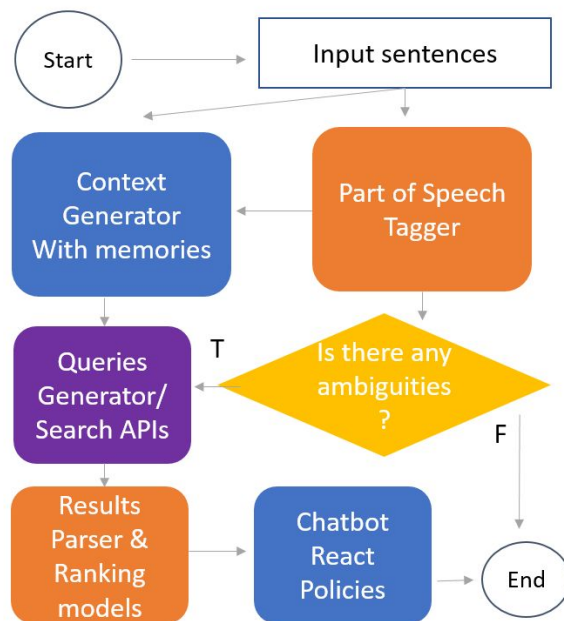


NLP techniques are very popular these days. One of its most interesting applications is building an interactive intelligent chatbot that could serve as a personal assistant. There is a problem with current chatbots: they could not resolve semantic ambiguities. As the picture shows, the word 'Barcelona' has at least two meanings: a football club and a city which belongs to Catalonia district. The dialogue in the figure happened on the day when Catalonia declared its independence. We think the chatbot should know we are not talking about the soccer team according to the conversation before. This project could lead to personal interactive news assistant and more intelligent chatbot.

Solution Description

By reading the context and googling the word 'Barcelona' at the same time, it should resolve the problem happened in the last section. Thus, a solution should involve a POS tagger which can detect nouns, a context generator and storage to extract context features and dynamically updated its memory of conversation topics, a query generator to generate queries and fetch results from google search API, a ranking models to pick the most relevant semantic meanings given the context features and reaction policies of the chatbot.

The architecture/workflow of our solution is shown below.



Milestones and Deliverables

Timeline	Goals
Week of 6th Nov	1.Design Architecture and algorithm; 2. Finish NLP models and chatbot ;3. Done with Project management logistics;4. Create Web crawler prototype
Week of 13th Nov	1.Test baseline prototype. ;2. Improve semantic ambiguities detection and results ranking models; 3. Design demo and test cases
Week of 20th Nov	1. Finish second iteration, minimal deliverables reached. ;2. Improve reacting policies;3. Start engineering with chatbot.
Week of 27th Nov	1. Fine tuning the model. ; 2. Finish engineering stuff.
Week of 4th Dec	1.Finish documentations and demo

Our primary goal and deliverable is a function that could decide the best semantic meanings for a noun given the context/conversation history. We will focus on algorithms in the first two weeks then try to integrate this function into some chatbot system.