CS410 Project Proposal

Karthik Venkat Ramanan

kv16@illinois.edu

1. What are the names and NetIDs of all your team members? Who is the captain?
My netid is kv16. I'm the only member of my team. My team name is "peak".

2. What topic have you chosen?
In my project, I aim to develop a chrome extension for easier search. I'd like to address the specific problem where, while browsing a website say a Youtube comment section, we encounter a word or phrase that we'd like to look up. This involves copy-pasting the word to a search engine and navigating to the article we want, which is often quite low on the search results because it is missing the specific context from the website we were browsing. My extension will use the local context of the word/phrase as well as the context of the web page to perform entity linking. It will then display the results of the lookup from a Knowledge Base / Search Engine in a window on the original web page without the user having to open a new tab. Another goal of mine is to encourage "linked thinking". My browser extension will also display neighboring entities and relations from the knowledge graph from which the entity is extracted and also use KG embeddings to perform a nearest-neighbor search to identify related entities.

3. Briefly describe any datasets, algorithms, or techniques you plan to use?
I plan on using statistical IR techniques like BM25 and PL2. I'll be using knowledge bases like Wikidata and DBpedia to extract relevant information. I'll also be using KG embedding techniques like ComplEx, TransE, etc.

4. How will you demonstrate that your approach will work as expected?
By delivering a working browser extension.

5. Which programming language do you plan to use?
The back-end will be in Python using Django. The front end will be in Javascript and Bootstrap.

6. Please justify that the workload of your topic is at least 20*N hours
The core tasks to be completed are-

Task	Time Commitment
Implementing entity linking on the back-end	10 hours
Implementing the front-end	10 hours

Displaying graph data from the KG	10 hours
KG nearest neighbors search	10 hours