Methodology:

A scraper was developed to extract information. This scraper was applied to the social media account of a user, who granted us access to their account. The information gathered included the biography section of their profile page. This information contained degrees, schools, hometown, pronunciation of their name, and where the user currently lives in. Additional information collected included the hobbies and the user’s friends. All this is done through the tags Facebook uses to identify the information. Lastly, the posts made by the user are parsed and stored at first to be later processed.

Once all the information from the users account has been scraped, different entities are extracted and identified. This is accomplished using Natural Language Toolkit (NLTK) to determine the different parts of speech in the information that was retrieved from Facebook. During this entity extraction, the words in every sentence are classified (adjectives, nouns, verbs, pronouns…) and relationships between the words are determined. For instance, if a user named Tim posts “Just had donuts for the first time, they were great”, the words “donuts” and “great” would tell us what Tim thinks about donuts.

The next step is to model all this data into a knowledge graph using ontologies. The ontologies used to describe the relationships between users and the data retrieved from Facebook are Friend Of A Friend (FOAF) and Semantically-Interlinked Online Communities (SIOC). FOAF is an existing ontology that describes persons, their activities and relations to other persons and objects. SIOC is another existing ontology used to describe a person’s information stored in internet discussion methods such as blogs, forums, and mailing list among others. All the information sorted in the ontologies is to be added to knowledge graphs. In the knowledge graph, each person that has any relationship with the user is a different node. After this, the user hobbies along with the users’ basic information were also added to the graph to populate the main user node.

Diagram

Description automatically generated

Figure 1. Data Flow Diagram

This diagram shows the data flow of the system. As explained in the previous paragraphs, the information comes from Facebook and then is parsed through the tag and post information, which will later be used to extract the entities and processed to create the ontology with the different relationships those entities have. Finally, the ontology will be used to generate the nodes that will be used to populate the knowledge graph.