IRE Assignment 2

Name: Aryan Jain
Roll Num: 2019101056

Wikidata :

Wikidata is a knowledge database, which contains a millions of statements. Some of them include like "*Narendra Modi is the prime minister of India.*", "The capital of Japan is Tokyo." or "Water has a boiling point of 100 degree Celsius." etc. Wikidata is a collaboratively edited multilingual knowledge graph hosted by the Wikimedia Foundation. It is a common source of open data that Wikimedia projects such as Wikipedia, and anyone else, can use. Wikidata is a free and open knowledge base that can be read and edited by both humans and machines. Wikidata acts as central storage for the **structured data** of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wiktionary, Wikisource, and others.

SPARQL :

SPARQL is a RDF(Resource Description Framework) query language i.e., a semantic query language for database, which is able to retrieve and manipulate data stored in RDF format. It is basically a language to formulate questions (queries) for knowledge databases. For eg. *"Who is the prime minister of India?", "What is the capital of Japan?", "What is the boiling point of the water?" etc.*
Having been designed to query a great variety of data, it can efficiently extract information hidden in non-uniform data and stored in various formats and sources.

Approach :

Firstly, i have gone through all the links given in the assignment as a pre-requisite to gain decent knowledge about what the Wikidata and SPARQL is, and how they works. (also re-looked up at the class for the same)
After having a good hand on the SPARQL queries, i jumped into the coding part. I have picked my topic as "***Narendra Modi***", for my assignment. Wrote queries to extract information in ***hindi***(my native language) language. I have created respective template statement for the results of the queries to fit into the statement in hindi. For some of the queries, hindi version of the query isn't available, so we need to leave those statements while creating the wikipage(just ignore those sentences). The hindi text for each query is shown on the terminal as well as stored in '***hindi_lang.txt***' file.
The code is written in such a manner, that it would also work for any other prime minister of any other country also, for whom corresponding information is also available on wikidata like "***Joe Biden" or "Imran Khan"***. For that we just need to update the variable "***pm_id***" with the QID of the respective person.
Appropriate comments are added for better understanding of the code.
After we have got all the sentences by running the python script, i have manually picked good hindi sentences and copied them in wiki-sandbox, and did the editing.

Sandbox Link :- *https://en.wikipedia.org/wiki/User:Rockarya2001/sandbox*
* in case above link doesn't work, then i also have attached screenshot for the same.