**Hyperlink Retrieval**

Each Wikipedia article has a number of hyperlinks. Sometimes a user may need to know what are all the other Wikipedia articles that are linked to word he searched so that they can have complete understanding of the topic. Going over each and every hyperlink line by line is a cumbersome task. So this program gets the input word and displays all the Wikipedia links associated with it. In addition to that, we implement two feature in this program – Relational searches based on Semantics and Recommendations based on the Hyperlink-Graph model. The program runs as follows: first the user has to enter the word whose related articles that he wishes to see. Then program gets the input word and visits the respective Wikipedia page and scraps its source code. Then a simple regex is ran to filter only the attributes of the HTML code. To the above results, a regex is applied to filter only the Wikipedia links and display the results on the screen. Now we have to find the words that conforms to its semantics. The search algorithms that we have developed identify the fragment within the database by using the relationships/links between the objects and their types in the fragment, and by attempting to find a “similar” pattern in the database. That is, to find words are of same meaning. Then wiki links of each word is fetched. Next, the program examines the search queries of previous users. It recommends interest items for users that is found from others that have similar preferences. This is established by constructing a directed Web graph and build its adjacent matrix with the edges of the graph being the hyperlink, and the vertices of the graph being the user. Now the user can click the link that he wants to read and it is redirected to the respective page making the navigation process simple. So this program makes it simple for the user to read all the related Wikipedia articles by improving its search and recommendation feature. This program is implemented in the python language.

Module Description

1)MODULE NAME: Pre processing

INPUT: Word

OUTPUT: Source Code

DESCRIPTION: The url for Wikipedia page is generated from search word and the source code is fetched

2)MODULE NAME: Tokenization

INPUT: Word

OUTPUT: Dictionary object

DESCRIPTION: The local file is accessed via shelve object and new key for the word is created

3)MODULE NAME: URL Extraction

INPUT: source code

OUTPUT: hyperlink

DESCRIPTION: Regular expression to filter out wiki links is applied on the source code. Then unnecessary links and duplicates are removed

4)MODULE NAME: Database Storage

INPUT: hyperlinks

OUTPUT: local file

DESCRIPTION: The hyperlinks are assigned to key of dictionary. It is stored to local file via shelve object

5)MODULE NAME: Finding related terms

INPUT: word

OUTPUT: synonyms

DESCRIPTION: The search word is passed to WordNet which in turn returns instance of it. Then similar words are matched.

6)MODULE NAME: Comparison

INPUT: word

OUTPUT: similar terms

DESCRIPTION: The search term is passed to find the degree of semantic relationship between it, the word from local file and its hyperlinks .

7)MODULE NAME:Conclusion

INPUT: Words

OUTPUT: Display Result

DESCRIPTION: Display the words matching to the search term semanticaly

**Project Done By:**

G.Gowshik Ajith (2016103525)

J.Silambarasan (2016103592)