|  |
| --- |
| **The University of Texas at Dallas**  **CS 6322**  **Information Retrieval**  **Spring 2021**    **Class Project Proposal**  Project TITLE: Search Engine for Startups  Group: 003    Students: Abdelmounaim Hafid, [axh170730@utdallas.edu](mailto:axh170730@utdallas.edu)  Abhishek Thurlapati, vxt160630@utdallas.edu  Dat Quốc Ngo, dqn170000@utdallas.edu |

1. The Team

…………………………………………………………………………..

In this project, our team built a search engine for ‘Startups’. We commit to have the following distribution of work and collaboration:

**Abhishek Thurlapati** shall be responsible for crawling the web to assemble our collection of web pages and its Web graph. A collection of 100,000 web pages shall be crawled. This student will deliver the results of the incremental crawl to the student responsible for indexing and relevance models. The architecture of the crawler shall be presented in our team’s project reports and discussions of its functionality shall be detailed.

**Abdelmounaim Hafid** shall be responsible for incremental indexing the crawled web pages and for the link analysis of the Web graph that was retrieved by crawling. The same student shall create two relevance models: (1) vector space relevance model as well as (2) relevance models based on PageRanking and HITS (and their combinations with the vector-based relevance models). This student shall collaborate with Niklas Borcanski (to import the crawled web pages) and with Servando Luviano to display the relevance results against any query as well as with Abhishek Thurlapati (to include clustering information in each of the relevance models) and with Dat Quốc Ngo to showcase the relevance results obtained for query expansion.

**Dat Quoc Ngo** shall be responsible for enabling the queries to be entered and the results to be presented in a graphical user interface embedded in the web page that hosts your search engine. When a query is entered, the search engine should present in separate search engine web page frames of the web page the following:

* Results of your relevance models against the query – all of those that you have implemented;
* Results of the search engine relevance against the query when clustering has been incorporated – you should present the clusters that you have obtained as well;
* Results of the search engine against the query when query expansion has been enabled – you should present the expanded query as well; • Results of Google against the query;
* Results of Bing against the query.

This student shall collaborate with Abdelmounaim Hafid (to provide the query to the relevance models and obtain the relevance results) and with Abhishek Thurlapati to obtain the clusters of the Web collection as well as the results of using the clusters in the relevance models. In addition, collaboration with student Dat Quốc Ngo is required to obtain the expanded queries and the corresponding relevance results.

All students are responsible for the creation of the search engine –that should run correctly during the presentation of the project.