Project Proposal

Chrome Extension – Search Campuswire and Coursera

# Team Details

I will be the only team member for this project. I am solely responsible for planning, executing and delivering the project work.

Team Member: Monika Thotha

NetId: mthotha2

# Topic

**Theme: Intelligent Browsing**

**Chrome Extension – Search Campuswire and Coursera.**

The topic I have chosen is Intelligent Browsing. I will be building a Chrome browser extension that will scrape through and index webpages of Campuswire and Coursera. The extension will provide user top 5 (or more) relevant documents from Coursera and/or Campuswire posts, each post will link to those documents.

The extension will provide unified results from Campuswire and/or Courser to users searching for a topic.

# Approach - Dataset, Algorithms and Techniques

At a high level the project can be broken down into following Milestones/Tasks:

Preprocessing:

1. Dataset: Write a crawler to get 100 (or more) Campuswire and 100 Coursera URLs.
2. Scrape the content of the pages.
   1. Algo: beautifulSoup
3. Data cleanup:
   1. Remove special characters, stopwords (Algo using NLTK).

Run-time activity:

1. Use ranking algorithm BM25 to calculate rank the relevant documents and select top 3.
2. Show first 50 (or less) characters of the document in the UI, along with the link to the document.

UI Design:

1. Write JS code to create a chrome extension to design UI for user to enter search query and show results.

The milestones and tasks shown above show how the search extension will combine the documents from two systems to provide user relevant documents.

The extension will provide a search box and a button where the user will be able to able to retrieve the relevant search results snippet with hyperlink to relevant documents.

# Programming Language

* Python for writing crawler and scraping the webpages.
* JavaScript for UI design of chrome extension.

# Task and time needed/Plan and Milestones

|  |  |  |
| --- | --- | --- |
| Milestone | Task | Hours |
| Dataset | Write crawlers and collect URLs | 5 |
|  | Write scrapers to get inner HTML | 2 |
|  | Data cleanup | 2 |
| Analyzing Search Query | Use BM25 to rank search results. | 5 |
| UI Design | Design chrome extension | 6 |
| Total Time | | 20 Hrs |

# Git Repo Link

<https://github.com/monikathotha/course-extension>