

# IR Assignment 5 Proposal for Usecase 1

Aakash Khadka, Gowtham Premkumar, Max Neubauer

In order to solve assignment 5 we implement a desktop application using python and the PyQt framework. The topics searchable by the user are related to computer science and machine learning only. The topics are stored in topics.txt. In order to be able to search with queries you first have to press a button to download the related documents from Wikipedia. After downloading the user is able to search. Searching is done using pylucene. The 10 most relevant documents are then displayed on the UI. The user is able to click a button to present the documents as clusters and vice versa. Clustering is done using scikit-learn and displayed by using matplotlib.

Technologies:

- python3
- PyQt ... for desktop app
- pylucene ... for searching
- scikit-learn ... for clustering
- matplotlib ... for plotting
- pandas ... for storing data

Data available / not available Load Data

Search query Search

Document 1

Document 2

Document 3

Document 4

Cluster

Data available / not available Load Data

Search query Search

Cluster image

show documents