# DESIGN DOCUMENT

Problem Statement : To be able to classify a web page from a URL and return a list of relevant topics.

Approach :

1. Accept the input URL by the user (via command line or via prompts) along with the maximum relevant keywords to display.
2. Get the HTML Document using the Jsoup library using Jsoup.connect(String url) method.
3. Parse the stop words and the candidate words using various validations such as checking if the word is a valid ASCII word, removing special characters from the word etc.
4. Store the candidate words along with frequency in a hash map and sort it in descending order of values (frequency of the keywords in the page).
5. Display the relevant topics as a group of keywords with same frequency together (with maximum frequency displayed the topmost).

Design Concepts Used :

1. Singleton Design Pattern for the 3 Classes.
2. Parser Interface implemented by HTML and Words Parser.
3. Setters and getters where essential.

Code is commented extensively in order to be understandable.
Exception Handling - Kept in mind from the beginning such as in case of handling any form of URL until it is valid.

Mayuresh Jakhotia