Team Members: What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

Archna Sobti (archnas2@illinois.edu) - Project Coordinator / Captain
Anjali Kumar (akumar4@illinois.edu)
Aryan Vakharia (aryannv2@illinois.edu)
Manu Pillai (mpilla6@illinois.edu)

Github:
https://github.com/Archna-1/CS410-final-project

Topic: What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

We chose a topic within intelligent browsing for our project: a chrome extension that summarizes news articles the user is viewing.

Most news articles are so long that a reader may be intimidated by the sheer information being presented. Having a tool that summarizes articles will make reading the news more accessible to everyone, while saving the user's time as well.

This relates to the theme of Text Information Systems and retrieval, as it involves summarizing and parsing text to contextualize information for the end user. It also serves as a practical application of knowledge derived from the coursework.

Techniques & Technical Details: Briefly describe any datasets, algorithms or techniques you plan to use.

We will be using typical NLP datasets like "stopwords.txt" (used in our MPs) to base our vocabulary of common words for our summary text retrieval system. Along with this, we will use an inverted index to store the highest ranked words for our retrieval system and the BM25 algorithm to enable fast search of such words.

Analysis: How will you demonstrate that your approach will work as expected?

We will verify that our approach will work as expected when we load up a random, lengthy article, open our extension and then find a result that closely matches the gist of the article.

Programming Language: Which programming language do you plan to use?

We plan to use: Python with MeTA toolkit, and JavaScript (Node.js + React.js) for the extension frontend and backend.

Time Commitment & Justification: Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

The workload of the topic is at least that much, since we plan to accomplish the following tasks:

1. Step 1: Project Planning (8-12 hours)
    a. Define the project scope and objectives (2-3 hours).
    b. Identify your target audience and their needs (2-3 hours).
    c. Set goals, milestones, and a timeline (2-3 hours).
    d. Research existing summarization techniques and technologies (2-3 hours).
2. Step 2: Design and Architecture (16-24 hours)
    a. Create mockups for the extension interface (4-6 hours).
    b. Plan the user experience and user interface design (4-6 hours).
    c. Define the data flow of the extension (4-6 hours).
3. Step 3: Development (32-48 hours)
    a. Set up the development environment (4-6 hours).
    b. Develop the core architecture of your Chrome extension (4-6 hours).
    c. Implement a content script for interacting with web pages (4-6 hours).
    d. Integrate a summarization algorithm (8-12 hours).
    e. Develop the user interface for the extension using React.js (8-12 hours).
4. Step 4: Testing and Debugging (12-18 hours)
    a. Thoroughly test the extension on various news websites and browsers (4-6 hours).
    b. Test the inverted index and BM25 algorithms with @NDCGk and F1 metrics
    c. Debug and optimize the extension's performance (4-6 hours).
    d. Ensure the extension works as expected with different article formats (4-6 hours).
5. Step 5: Deployment, Documentation, and Maintenance (12-18 hours)
    a. Attempt to prepare the extension for deployment on the Chrome Web Store (4-6 hours).
    b. Create user documentation, installation instructions, and FAQs (4-6 hours).
    c. Develop a plan for ongoing maintenance and updates (4-6 hours).