**Fall - 2024**

**Assignment 1**

* Due on Friday September 27th.
* Write your answer in this document below each exercise.
* Please write your name and student ID at the top of this word document

1. Search engine
2. Information extraction (parser)
3. **Information Extraction (Parser) Project**

**Design:** This program is designed to help users reduce screen time while browsing website headlines. It will retrieve and web scrape headlines from selected websites, displaying them as a list of clickable links. The links can be sorted by criteria such as total views or user ratings, allowing users to quickly access the most relevant or popular content.

**Development:** The project will be developed using Python within an Integrated Development Environment (IDE), specifically **Visual Studio**. The program will feature a series of graphical user interfaces (GUIs) to guide the user through the process. Once development is complete, the Python (.py) file will be converted into an executable (.exe) file. This .exe file can be easily shared through platforms like email or GitHub, eliminating the need for users to have Python or Visual Studio installed. They can simply double-click the .exe file from their desktop to run the program.

**GUI Phase:**

* The GUI will prompt the user to select a category of interest, such as **Sports** or **News**.
* Based on the selected category, the program will scrape and retrieve headlines from a corresponding website.
* The GUI will present the headlines in a sortable list format. Users can sort by factors like views or ratings, and clicking on any headline will open the associated link.

This design aims to simplify the process of browsing headlines, saving users time, and reducing their overall screen exposure.

1. Information retrieval
2. Clustering