

Weather Data

My task consists of scraping the National Weather Service (NWS) website for data regarding weather in cities and regions across the United States and being able to output relevant results based on a search. For example, if I make a query about a city, such as Omaha, a list of documents is returned with relevant results regarding NWS data for Omaha. If a query is made about snow, documents related to weather statements or pages regarding snow will be returned.

Essentially, this is a search engine for the NWS website. The website doesn't have a particularly robust search system in place and doesn't show up in Google results as readily as other weather websites (which typically don't have anywhere near the breadth of data NWS has).

Usage of NWS API is free for the public. It is rate limited but allows a generous number of requests to accommodate most usage, so there should be plenty of opportunity to use the data as needed.

This search engine could be expanded to include other climate-related government websites, such as those of the Natural Resources Conservation Service or the Environmental Protection Agency.

The search engine can also be tailored to give more specific results such as charts of weather data comparisons for different cities, but this may be more in the territory of database retrieval rather than text retrieval as defined in CS 410.

I expect I will discover some potential applications of concepts from class to be applied as I begin working on the project as well.

This project will be evaluated by measuring the effectiveness of the search engine. Have documents been collected effectively? Are the results ranked properly given a query?

I plan to achieve the task mainly using Python. I expect getting familiar with the API, creating the framework, and designing the query application and the ranking function for results will take well beyond 20 hours of work.