

Report on Assignment 6

For this assignment, I created a knowledge base regarding the San Francisco Bay Area. With a given starting URL, the script scrapes data from various sites. Using the metric tf-idf, the script iterates through the corpus and calculates a score for each token. These scores are used to generate a list of top terms. I then reduce the list by manually selecting the top 10 terms — [California, war, town, museum, Francisco, Bay, downtown, shipyards, data, Alviso]. Given this new list, the script creates a dictionary with entries (key: array of string facts). Iterating through the cleaned data, the script checks for the term's presence before adding facts to relevant keys

Sample of Knowledge Base:

Three Facts about California:

- Region of California in the United States San Francisco Bay Area Region of California Clockwise from top: The Golden Gate Bridge , Muir Woods National Monument , Napa Valley vineyards, Sather Tower at UC Berkeley with the Bay in the background, the Stanford University Oval, the Palo Alto Baylands Nature Preserve , downtown San Jose 's skyline, and San Francisco Chinatown with the Bay Bridge in the background Location of the Bay Area within California .

- Home to approximately 7.76 million people, Northern California 's nine-county Bay Area contains many cities, towns, airports, and associated regional, state, and national parks, connected by a complex multimodal transportation network.

- The larger federal classification, the combined statistical area of the region which includes fourteen counties, is the second-largest in California (after the Greater Los Angeles area), the fifth-largest in the United States, and the 41st-largest urban area in the world with 9.67 million people.

Three Facts about war:

- During World War II , the Bay Area played a major role in America's war effort in the Asiatic-Pacific Theater , with the San Francisco Port of Embarkation , of which Fort Mason was one of fourteen installations and location of the headquarters, acting as a primary embarkation point for American forces.

- Economically, the post-war Bay Area saw huge growth in the financial and technology industries, creating an economy with a gross domestic product of over \$700 billion, and home to the third-highest concentration of Fortune 500 companies in the United States (as of 2018).

- After the United States joined World War II in 1941, the Bay Area became a major domestic military and naval hub, with large shipyards constructed in Sausalito and across the East Bay to build ships for the war effort.

Knowledge Base

Sample Dialog

Hello! I'm a chat bot. Feel free to ask me questions about SFBA.

Which museum is open right now?

The De Young Museum is open at this current time.

Which county is Alviso in?

Alviso is located in Santa Clara County.

How can I travel to downtown SF?

Take the BART to Embarcadero.