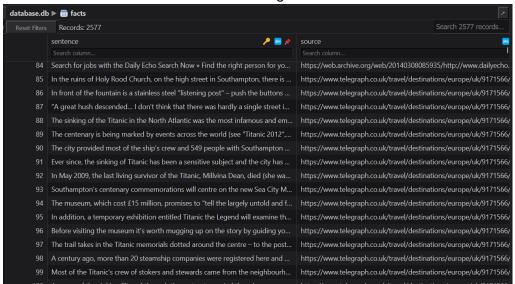
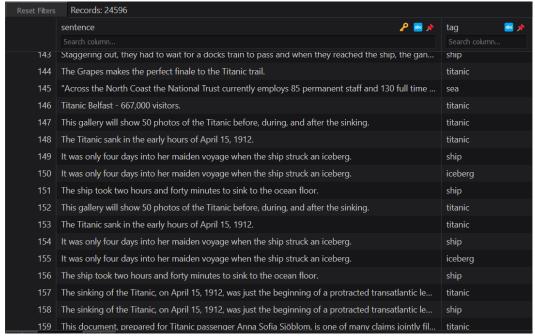
Web Crawler Discussion

Knowledge Base

A knowledge base generated by the web scraper code on the subject of the Titanic is implemented by using the sqlite3 Python library to create a simple relational database file. The database contains two tables: facts and tags.



Each entry in the facts table has a sentence and its source URL.



Each entry in the tags table has a foreign key to a sentence in the facts in the table and a tag string.

This system allows each fact entry to be tagged by multiple keywords since the sentence may be relevant to multiple subjects. Sentences were put in the database if they contained at least one of the tags. The tags used were 'titanic', 'ship', 'maritime', 'salvage', 'wreck', 'bodies', 'dead', 'iceberg', 'sea', and 'steel' which were among the 25 most common words across all scraped websites or were relevant to the topic. The scraping was limited to the tags to avoid website elements like menus.

Chatbot Sample Dialogue

Given below is an example dialogue for a chatbot using the information scraped from the Titanic webpage example.

RoseBot: Hello, my name is RoseBot, and I really like talking about the Titanic (the ship, not the movie). What's your name?

Jack (user): My name is Jack.

RoseBot: Hello, Jack. Is there anything you would like to know about the Titanic?

Jack (user): Yes, when did they spot the iceberg that hit the ship?

RoseBot: Good question! The iceberg was spotted on April 14th, 1912 at around 11:40 pm by the ship's local time. Would you have boarded the Titanic in 1912 if you had the chance?

Jack (user): That depends. How many people died?

RoseBot: 706 people survived and 1,517 people died.

Jack (user): I don't like those odds. No, I would not board the Titanic.

RoseBot: I would take my chances. I think I would survive by clinging to a piece of driftwood.