Omari Wilondja

Chaoqun Ni

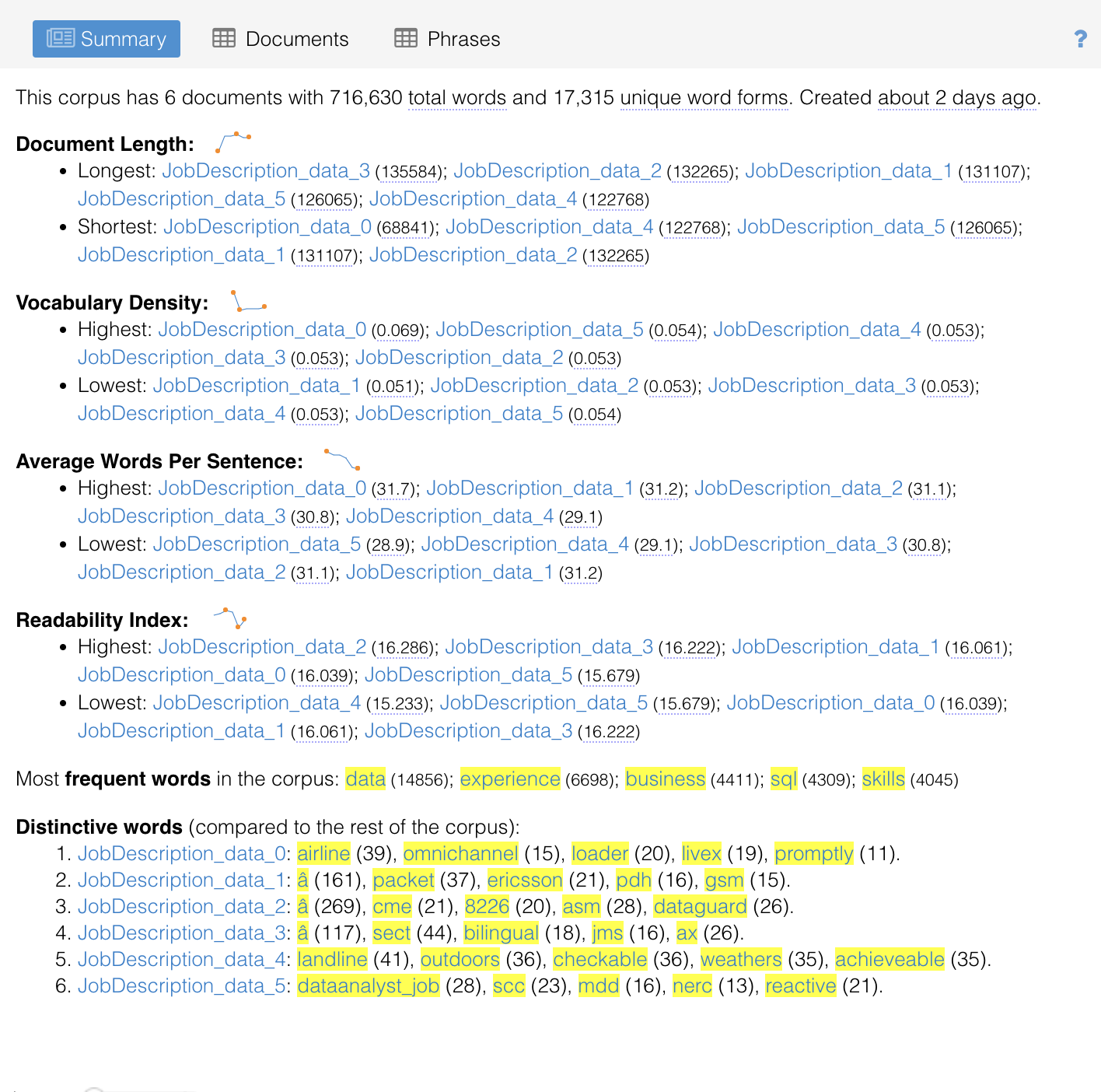
LIS 407

March 27, 2023

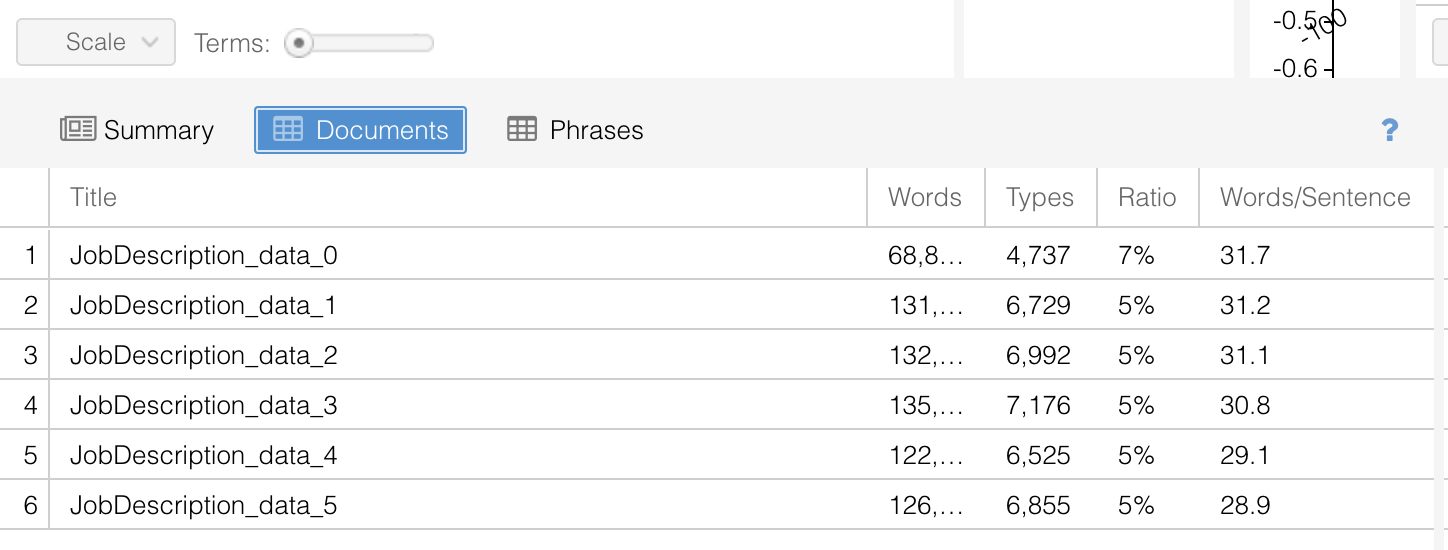
## Text Visualization

1. Provide a summary for each document, including the total number of words and the number of unique word forms in the whole collection. **[The lower-left corner of the result screen shows this.]**You may take a screenshot **or**choose to type it in.

**Answer**:



1. Provide the number of words, the number of word types, the ratio of words/tokens, and the number of words per sentence of each document. **[Click the Documents icon on the lower-left panel of the result screen.]**You may take a screenshot or choose to type it in.

**Answer: **

1. Provide a word cloud visualization of**20-40** words based on word frequency in the whole collection.**[Adjust the "Term" bar at the bottom of the word cloud on the upper left panel of the result screen.] Use one to two sentences to summarize the key message.**

**Answer:**

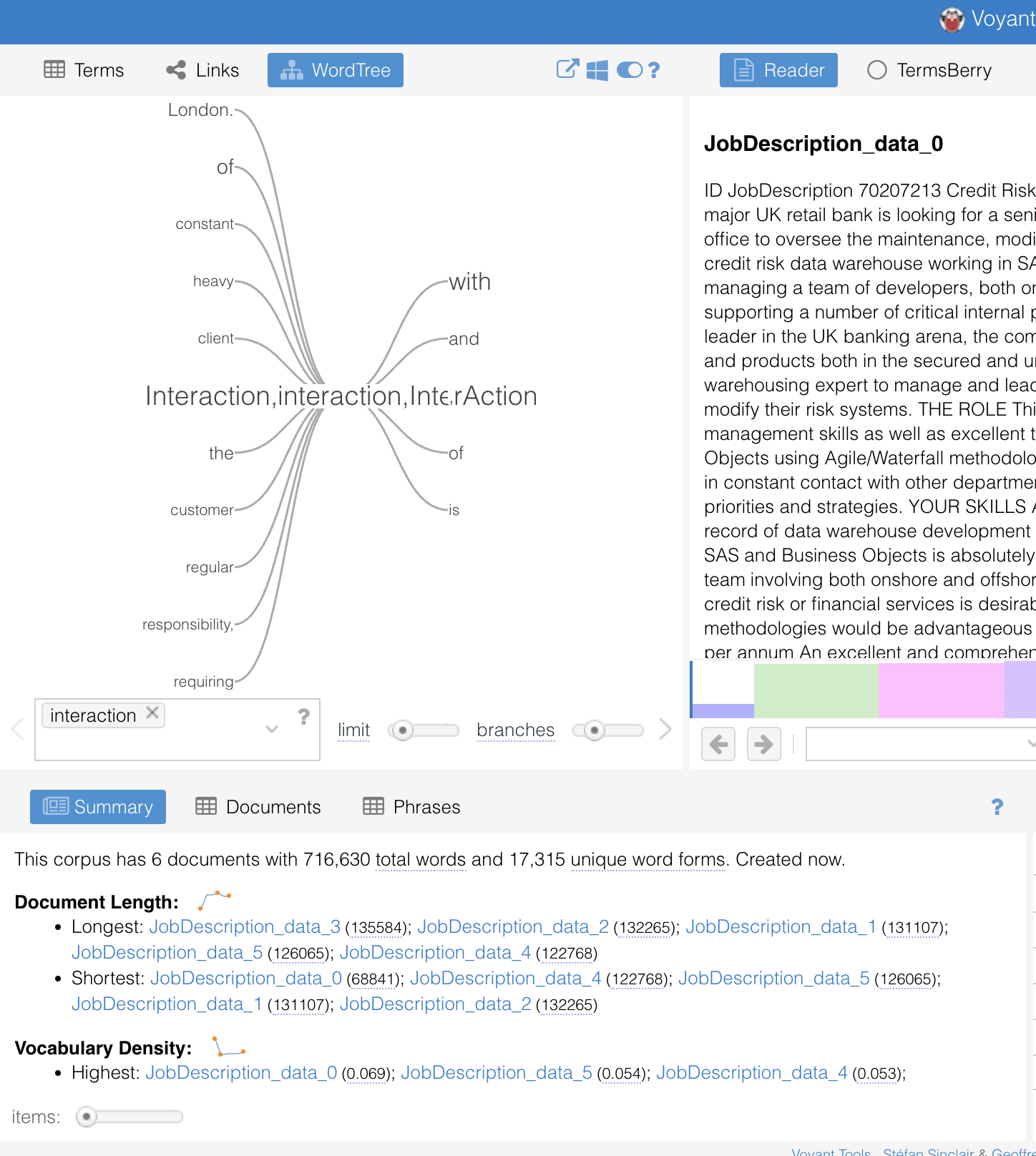
**Timeline

Description automatically generated**

* Overall, I believe this word cloud highlights some of the skills required to handle data or work with data or information in general.

1. Create a word tree based on the whole collection for the word "interaction" and provide the visualization. **[At the word cloud panel: click the question mark -> the windows sign in the middle->corpus tool->wordtree ]. Use one to two sentences to summarize the key message.**

Answer:



I noticed two interactions that ended in s and \*. I wasn't sure if we needed to use both for this word tree. Overall, the importance of communicating with customers and clients for business-related reasons is emphasized by this word tree.

1. Create a bar chart based on the number of times the word "data" appeared in each document. Each bar is a document.**Provide the visualizations in the assignment document. [The middle panel of the upper result screen. You need to choose your term at the bottom of the panel]. Use one to two sentences to summarize the key message.**

Answer:

Chart, bar chart

Description automatically generated

The essential message of the bar chart is the variation in the frequency of the term "data" across different documents in the corpus. Based on this image, it appears that document five had the highest frequency, followed by document six.

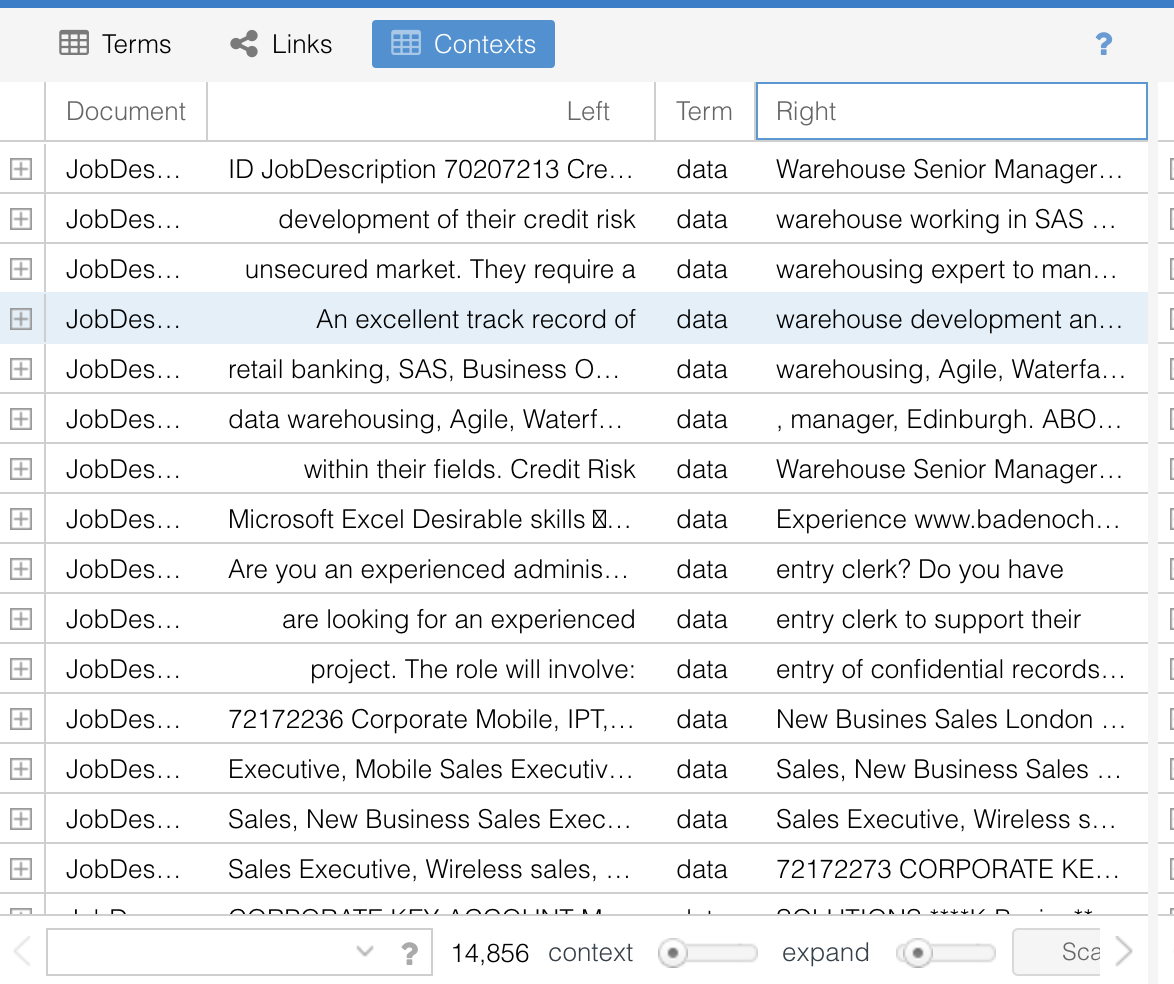
1. Create a line chart that shows the relative frequencies of "data", "SQL",  and "Python" in each document. Provide the visualization in the assignment document. **[The right panel of the upper result screen. You need to choose your terms at the bottom of the panel]. Use one to two sentences to summarize the key message.**

Chart, line chart

Description automatically generated

**The line chart's main message is the relative frequency of each term across different documents in the corpus. In every document, the word data was used more than python or sql. For example, it appears that sql was not widely utilized in most documents.**

1. Show the context for the keyword "data". Provide the visualization in the assignment document.**[The right panel of the bottom result screen. You need to add "data" as the keyword of interest at the bottom of the panel]. Use one to two sentences to summarize the key message.**



The visualization's main message is the surrounding text and context of the keyword "data" in each page in the corpus. Examining the context of the keyword "data", for example, I can use such method to understand how it is used and what it signifies in each document.

#### **Task 2: Preparing and Visualizing text from a single document**

Chart, bar chart

Description automatically generated

The bar chart displays the top 20 most frequent words in a text corpus, with the word "class" appearing 4000 times.

Text, whiteboard

Description automatically generated

The illustration of text data as a word cloud reveals that the word "class" appears to be more prominent, with a larger and clearer font size than other words.

Text

Description automatically generated

According to the bigram word cloud, the most prevalent pairs of words in the text corpus appear to be related to education and tests, such as teacher-student, professor-student, and final exam.