WEEK 5 RESEARCH NOTEBOOK UPDATE:

A research technique called content analysis examines textual or visual data to find patterns and themes. It is frequently employed in the social sciences, which encompasses disciplines including education, psychology, sociology, and communication studies.

A type of content analysis called theme analysis focuses especially on finding themes or patterns in the material. In other words, it entails grouping data according to themes or subjects that show up in the data.

Content and theme analysis is a potent research technique for identifying trends and themes in textual or visual data. To make sure that your analysis is accurate and valid, you need thoroughly identify your research topic and coding scheme.

The similarity between the rankings of terms in two or more papers is measured by rank similarity. The degree to which a certain document adheres to a user's search criteria is frequently determined using this metric in information retrieval. The goal is to find pages that contain words that are identical to or similar to those in the user's query and rank them according to their relevance.

On the other hand, document similarity gauges how similar the content of two or more documents is. This metric is frequently used in machine learning and natural language processing to gauge how similar distinct texts are to one another. Many techniques, like cosine similarity and Jaccard similarity, can be used to determine how similar two documents are.

In order to determine how similar two documents are, the cosine similarity method computes the angle between two vectors that reflect the frequency of phrases in each text. The scale for this measurement is 0 (no resemblance) to 1. (perfect similarity).

The intersection of the set of terms in each document divided by the union of the set of terms in each document is how the Jaccard similarity

metric determines how similar two documents are. The scale for this measurement is 0 (no resemblance) to 1. (perfect similarity).

In text analysis, rank similarity and document similarity are both helpful metrics that may be applied to a range of operations like topic modelling, document grouping, and information retrieval.

The above-mentioned methods would be a good fit for our project.