

NLTK Method

This method does not make use of OpenAI's text processing APIs and instead it makes use of Python's Natural Language Toolkit, also known as NLTK for short. In order to match up posts with other similar ones, we will need to generate a list of keywords for each posted question. To do this, we will take the input text, lemmatize and preprocess the resulting tokens (such as removing common "stop words" from the text), and then finally use the include Rapid Automatic Keyword Extraction (RAKE) algorithm to generate a list of keywords. This will be compared to the keywords of the other posts, and those that have a sufficient number of matching keywords will be recommended.

OpenAI Method

This method had the same goal as the NLTK Method, but instead makes use of OpenAI's ability to extract keywords from a text automatically. Just like before these keywords will be used to judge related texts. In addition, OpenAI's text embedding and clustering features allow for an extra method in which to judge how similar an 2 posts are. Once again, those that are closely clustered enough and have enough matching keywords will be recommended to the user.