

Identifying similarity between text is a common problem in NLP and is used by many companies world wide. The most common application of text similarity comes from the form of identifying plagiarized text. Educational facilities ranging from elementary school, high school, college and universities all around the world use services like Turnitin to ensure the work submitted by students is original and their own. Other applications of text similarity is commonly used by companies which have a similar structure to Stack Overflow or Stack Exchange. They want to be able to identify and flag duplicated questions so the user posting the question can be referenced to the original post with the solution. This increases the number of unique questions being asked on their platform.

Text similarity