Lecture 2 Project: Document Distance

Yanuar Heru Prakosa

31-05-2021

The Document Distance

The assignment from the Lecture 2 in 6-006 is related to one called document distance. Thus I need to find out first what is exactly document distance is. Form this MIT discussion webpage, I learn that document distance will be measured using vector equations. Let D be a text, then a word is a consecutive of alphanumeric characters. We will not distinguish upper case and lower case letters, but we use non alphanumeric characters as delimiter between words. For example the word "can't" consist of 2 words: can and t.