A Quick Introduction to Topic Modelling

HACKING THE NEWS WORKSHOP

Basic Description

"if you have hundreds of documents from an archive and you wish to understand something of what the archive contains without necessarily reading every document, then topic modeling might be a good approach."

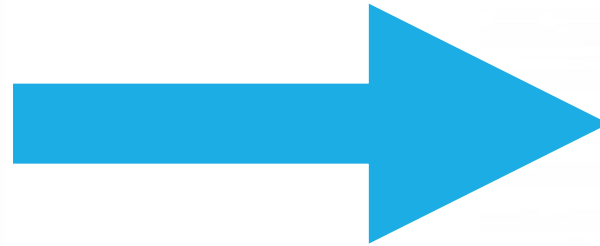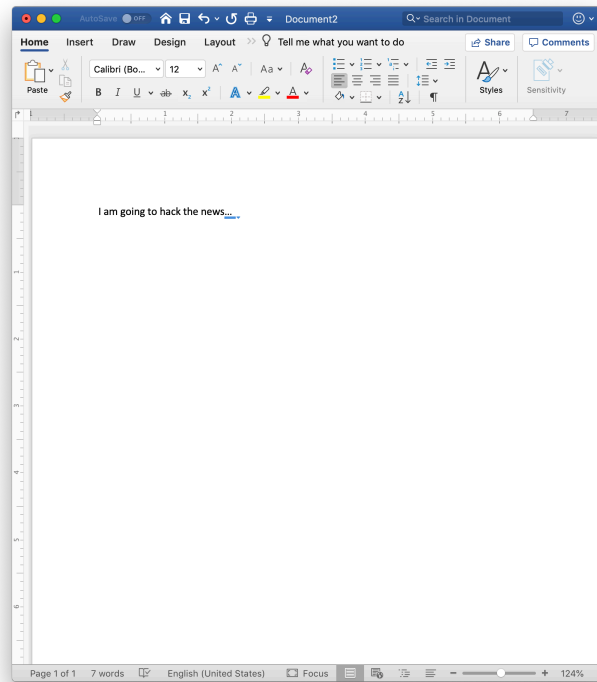# Takes data and finds "topics" in unstructured data

Description from "MALLET"

"Topic models provide a simple way to analyze large volumes of unlabeled text. A "topic" consists of a cluster of words that frequently occur together. Using contextual clues, topic models can connect words with similar meanings and distinguish between uses of words with multiple meanings."

ZZZZ

Idea: When I write, I pick words for topics from a big bag.

i.e.

when I'm writing about labour unions, I use certain words: "unions," "labour," "workers," "men," etc.

What if we could reverse that process?

If we had an entire workshop dedicated to this...

But we don't so let's use the topic modelling tool!