

Taxonomy Automation(Text classification of Reviews)

Nishant Bannikoppa
MSc in Big Data

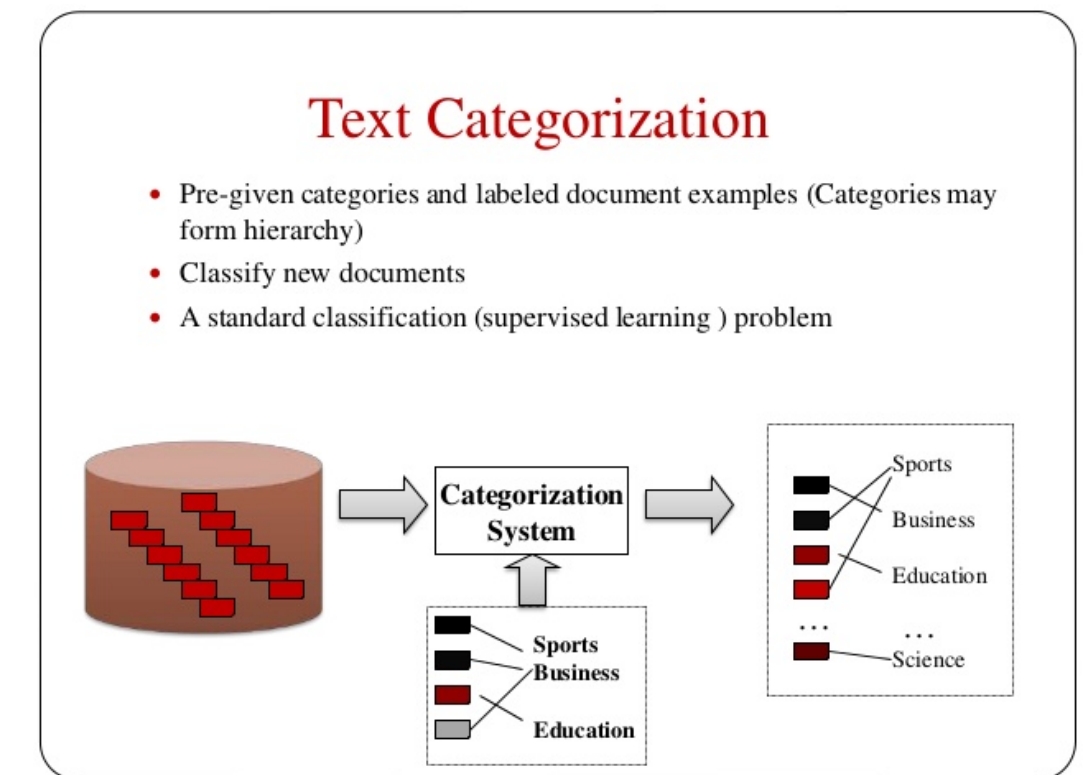
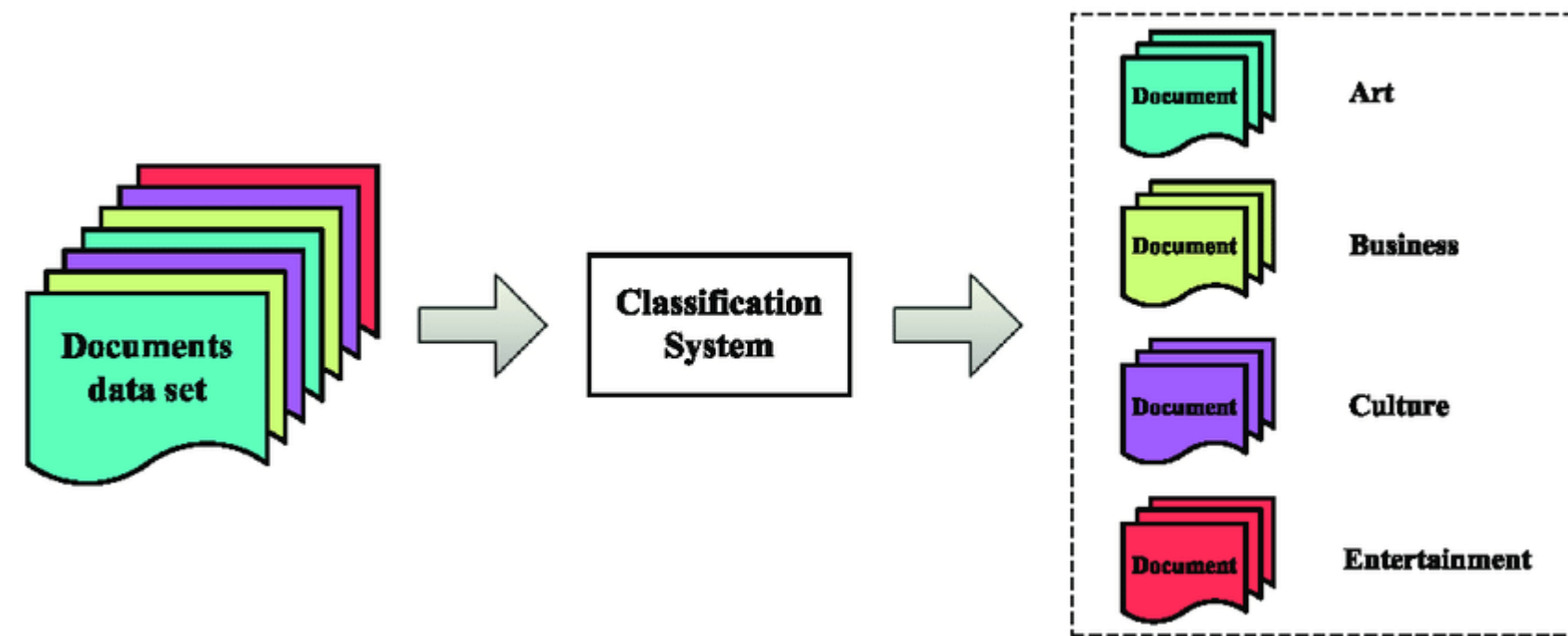
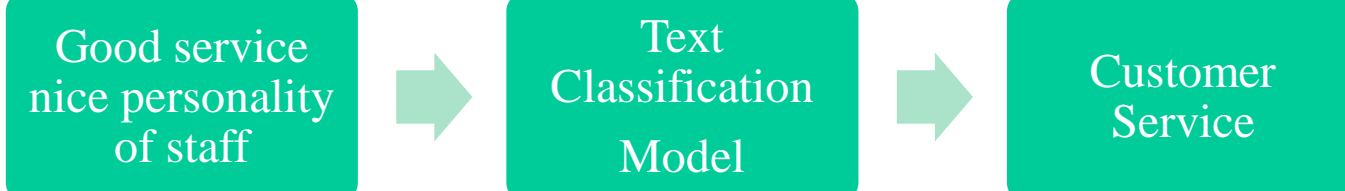
Abstract

What is Taxonomy Automation.

Taxonomy is a process of naming, describing and classifying the data into sets on the basis of their categories. Hence, taxonomy automation is creating a system which automatically classifies the data into their categories without much human intervention.

Background

The primary task of the project is the Classifying the reviews from the customers i.e., Text classification. Extracting data from messy text by analysing it's grammatical structure is very challenging because the text doesn't follow normal grammatical rules. Text classification can be done in two different ways: manual and automatic classification. In the former, a human annotator interprets the content of text and categorizes it accordingly. This method usually can provide quality results but it's time-consuming and expensive. The latter applies [machine learning](#), natural language processing, and other techniques to automatically classify text in a faster and more cost-effective way.



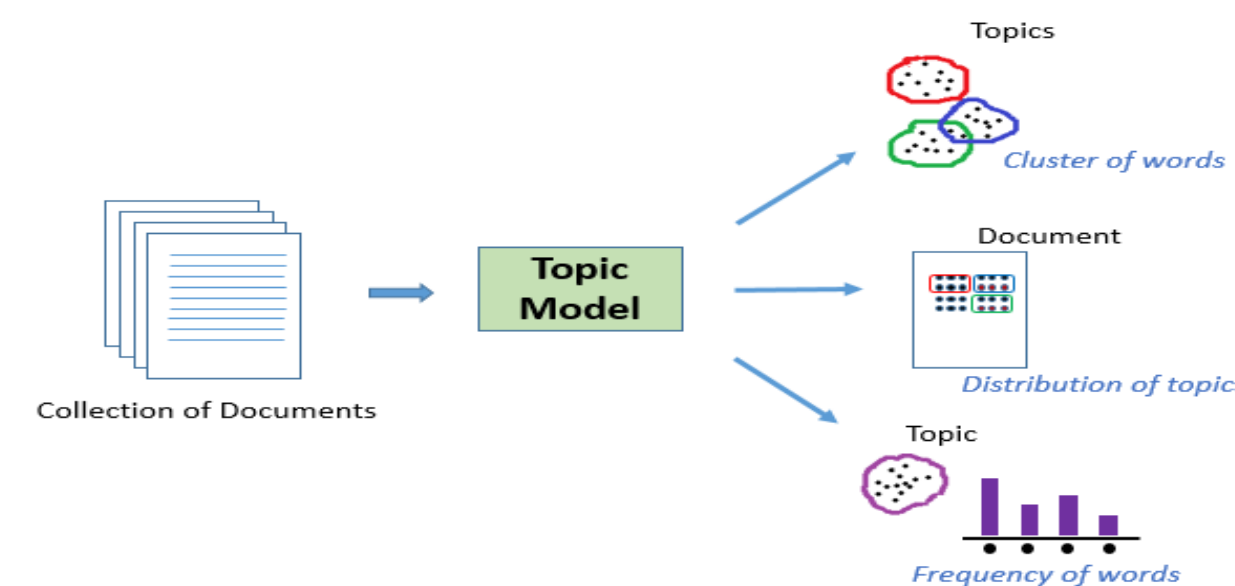
Methodologies

NLP

Nlp stands for Natural Language processing.

Major nlp methods used for text classification are

- Lemmitization.
- Sentence breaking.
- Word segmentation.
- Terminology extraction.
- Topic segmentation
- Relationship extraction
- Sentiment Analysis



Process

