**Text Analysis**

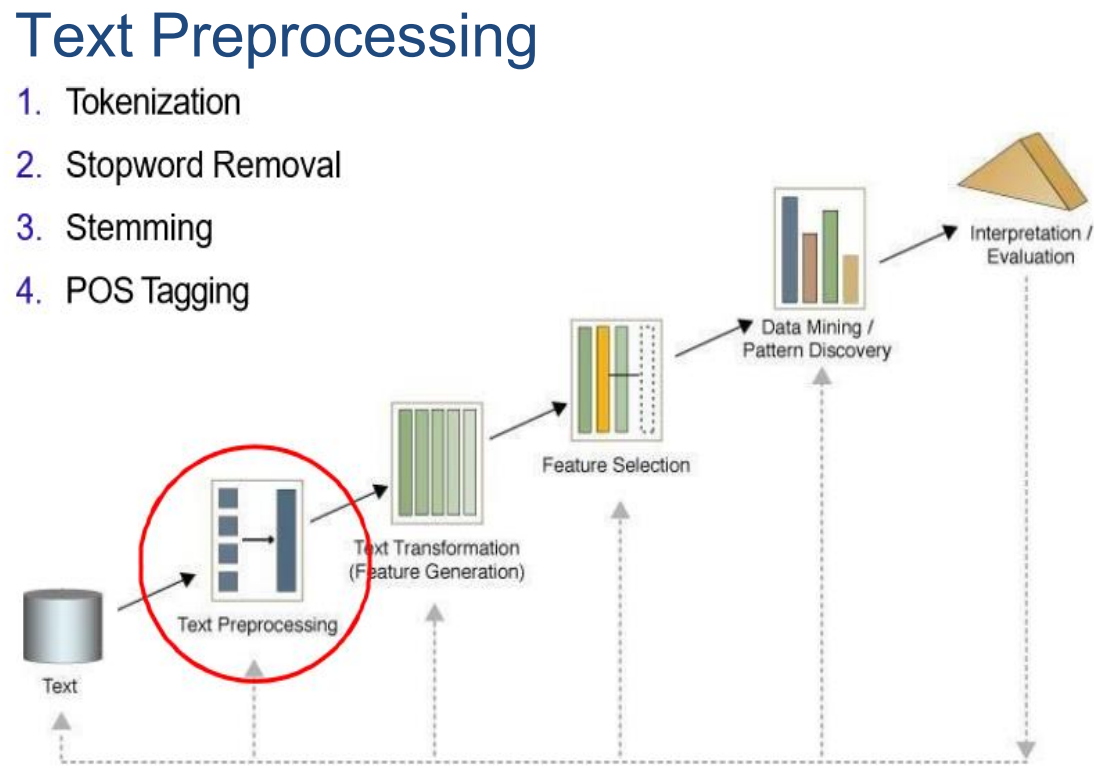
## **What is text analysis?**

Text analysis is the process of using computer systems to **read and understand human-written text for business insights**. Text analysis software can independently classify, sort, and extract information from text to identify **patterns, relationships**, **sentiments**, and other actionable **knowledge**. You can use text analysis to efficiently and accurately process multiple text-based sources such as **emails, documents, social media content, and product reviews, like a human would.**

## **Why is text analysis important?**

Businesses use text analysis to extract actionable insights from various unstructured data sources. **They depend on feedback from sources like emails, social media, and customer survey responses to aid decision-making.**

### 



## 

## **What are the stages in text analysis?**

To implement text analysis, you need to follow a systematic process that goes through four stages.

### **Stage 1—Data gathering**

### **Stage 2—Data preparation**

Data preparation is an essential part of text analysis. It involves structuring raw text data in an acceptable format for analysis. The text analysis software automates the process and involves the following common natural language processing (NLP) methods.

**Tokenization**

Tokenizing is the process of converting a sentence into a group of words. Tokenization is segregating the raw text into multiple parts that make semantic sense.

**Part-of-speech(POS) tagging**

Part-of-speech tagging assigns grammatical tags to the tokenized text. For example, applying this step to the previously mentioned tokens results in *text: Noun; analytics: Noun; benefits: Verb; businesses: Noun*.

**Stemming**

* It is the NLP task of removing inflection through **dropping unnecessary characters**

( wait-> waiting , waited, waits)

* Words are stripped to become stems
* The **output** base form **does not necessarily exist in a human language dictionary**.

**Lemmatization**

* **Lemmatization:** It is the NLP task of removing inflection by **finding the base form of text from the dictionary**.
* The output base form **always exist in the dictionary.**

Lemmatization is a linguistic process that simplifies words into their **dictionary form**.

For example, the dictionary form of *visualizing* is *visualize*.

**Stop words removal**

Stop words are words that offer little or no semantic context to a sentence, such as *and*, *or*, and *for*.

* **Sample uses of text processing:**

1. **Sentiment analysis** : is the most common **text classification tool** that **analysis** an incoming **message** and tells whether the underlying sentiment is **positive**, **negative** or **neutral**.
2. **Text Classification:** spam filtering
3. **Text clustering**

### 

### **Sentiment analysis**

Sentiment analysis or opinion mining uses text analysis methods to understand the opinion conveyed in a piece of text. You can use sentiment analysis of reviews, blogs, forums, and other online media to determine if your customers are happy with their purchases. By using sentiment analysis and identifying specific keywords, you can track changes in customer opinion and identify the root cause of the problem.

### **Text analysis vs. text analytics**

Text analytics helps you determine if there’s a particular trend or pattern from the results of analyzing thousands of pieces of feedback. Meanwhile, you can use text analysis to determine whether a customer’s feedback is positive or negative.

**References**

https://aws.amazon.com/what-is/text-analysis/