Great Place To Work. Certified MAR 2021–FEB 2022 INDIA

UPES

UNIVERSITY OF TOMORROW

nirf RANKING 2021
Ranked amongst **top 100** universities in **India**

Accredited **Grade** 'A' by NAAC

QS STARS RATING SYSTEM 2020
QS 5 Star Rating for Academic Development, Employability, Facilities and Program Strength

Perfect score of **150/150** as a testament to exceptional E-Learning methods

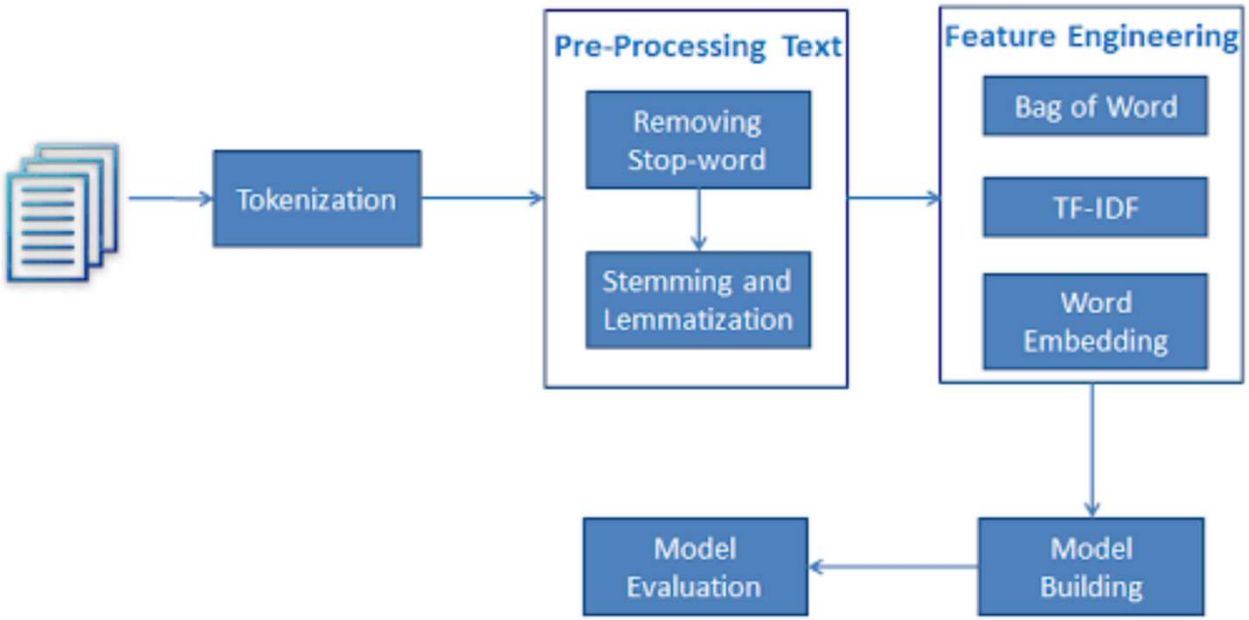
University of the Year (North India) awarded by ASSOCHAM

SAFE GUARD
Certified for **safety and hygiene** by Bureau Veritas

Semantic analysis

- Concept identify and extract the meaning specified according to a language dictionary.
- Semantic analyzer: Extraction of the meaningfulness.
- Identify the meaning of individual words.
- Identify the meaning of the sentence: Combining the meaning of the individual words.
- The purpose of semantic analysis is to draw exact meaning, or you can say dictionary meaning from the text. The work of semantic analyzer is to check the text for meaningfulness.

Elements of Semantic Analysis



Hyponymy: It may be defined as the relationship between a generic term and instances of that generic term. Here the generic term is called hypernym and its instances are called hyponyms. For example, the word color is hypernym and the color blue, yellow etc. are hyponyms

Elements of Semantic Analysis

Hyponymy: It may be defined as the relationship between a generic term and instances of that generic term. Here the generic term is called hypernym and its instances are called hyponyms. For example, the word color is hypernym and the color blue, yellow etc. are hyponyms.

Homonymy: It may be defined as the words having same spelling or same form but having different and unrelated meaning. For example, the word “Bat” is a homonymy word because bat can be an implement to hit a ball or bat is a nocturnal flying also. (Relationship between two or more words : similar spelling : different meaning).

Synonymy

It is the relation between two lexical items having different forms but expressing the same or a close meaning. Examples are ‘author/writer’, ‘fate/destiny’.

Antonymy: Relationship between two lexical items : provide dissimilarity in their meaning.

Building Blocks of Semantic System

In word representation or representation of the meaning of the words, the following building blocks play an important role:

Entities − It represents the individual such as a particular person, location etc. For example, Haryana. India, Ram all are entities.

Concepts − It represents the general category of the individuals such as a person, city, etc.

Relations − It represents the relationship between entities and concept. For example, Ram is a person.

Predicates − It represents the verb structures. For example, semantic roles and case grammar are the examples of predicates.

Approaches to Meaning Representations

Semantic analysis uses the following approaches for the representation of meaning −

- First order predicate logic (FOPL)
- Semantic Nets
- Conceptual dependency (CD)
- Rule-based architecture
- Case Grammar
- Conceptual Graphs

Need of Meaning Representations

Linking of linguistic elements to non-linguistic elements
Representing variety at lexical level

Corpus creation

What is a corpus?

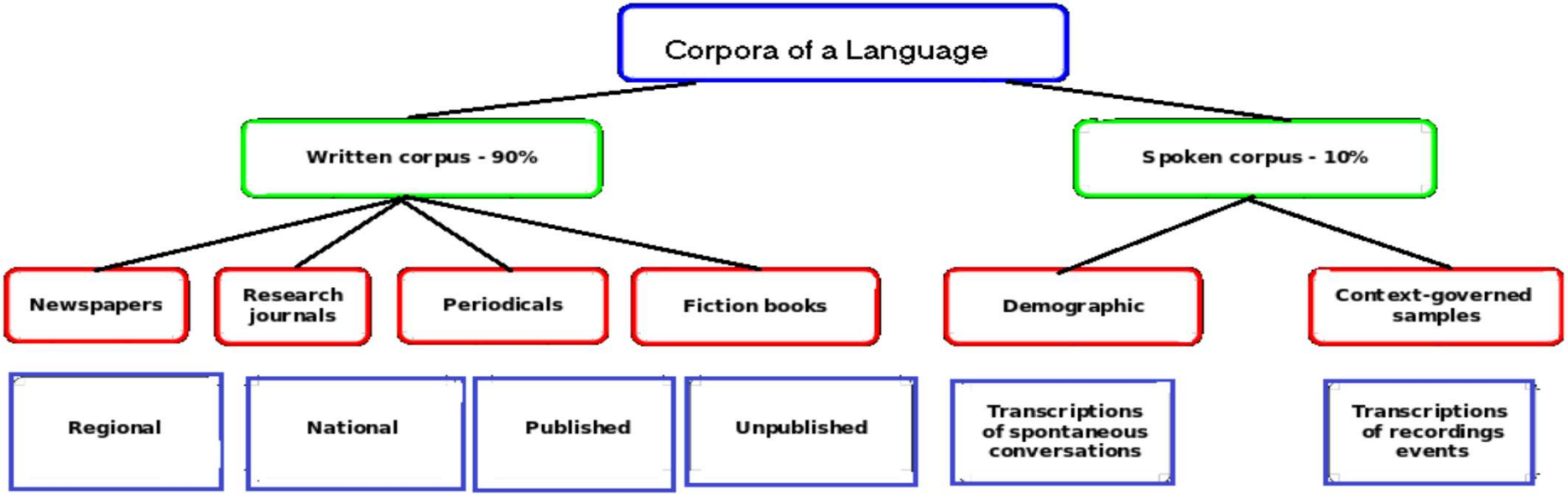
A corpus is a collection of authentic text or audio organized into datasets. Authentic means text written or audio spoken by a native of the language. A corpus can be made up of everything from newspapers, novels, recipes, radio broadcasts to television shows, movies, and tweets. In natural language processing, a corpus contains text and speech data that can be used to train AI and machine learning systems. If a user has a specific problem or objective they want to address, they'll need a collection of data that supports, or at least is a representation of, what they're looking to achieve with machine learning and NLP.

What are the features of a good corpus?

Large corpus size:
High-quality data:
Clean data:
Balance:

Corpus creation

- Corpus creation is **the process of building a dataset**.
- Basic reference location that the features of the natural language.
- Created based upon the grammar and context.
- Analysis of the Linguistics and Hypothesis testing.

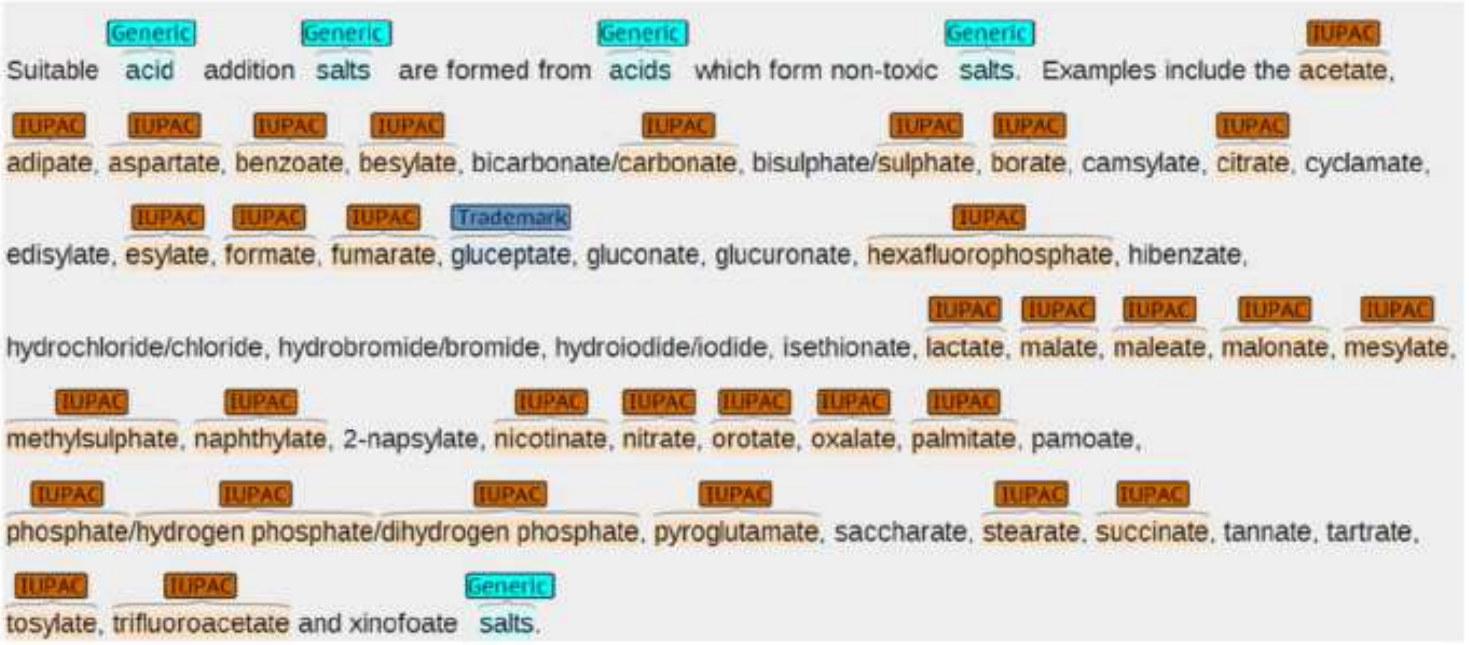


Corpus linguistics

- Collection of natural language elements that can characterize any natural language.
- Provide rules, grammar, and structure for the language.
- Corpus linguistics concepts:
  - Direct observation of the data.
  - Performance based upon the context.
  - Lexico grammatical approach.
  - Temporal activity based upon the context.
- Monolingual or Multilingual.
- Identify the state of any natural language.
- Teaching or research.
- Electronic Text Library (ETL).
- Holds a lot of text to identify patterns.

Annotations in text Corpus

Corpus annotation is the practice of adding interpretative linguistic information to a corpus. For example, one common type of annotation is the addition of tags, or labels, indicating the word class to which words in a text belong.



NLP task-specific training corpora

- POS Tagging:
  - Penn Treebank's WSJ section 45-tag tagset.
- Named entity recognition:
  - CoNLL 2003 NER task is newswire content from Reuters RCV1 corpus.
- Constituency parsing:
  - Penn Treebank's WSJ section has dataset.
- Semantic role labelling:
  - OntoNotes v5.0 with syntactic and semantic annotations.
- Sentiment analysis:
  - IMDb has released 50K movie reviews.
  - Amazon Customer Reviews of 130 million reviews
- Text Classification/Clustering:
  - Reuters-21578 news documents from 1987 indexed by categories.
- Question answering:
  - Jeopardy dataset of about 200000 Q & A

Areas using text annotations

- Question answering systems
  - Converse with the clients in any sort of environment wherever a question needs to be answered.
  - The questions and answers: Any natural language.
- Summarization
  - Provides very fast and elevated summary of very large documents in natural languages.
- Machine translation
  - The initial phases is NLP application of translation from one natural language to another.
- Speech recognition
  - Due to the automatic speech recognition systems, available most of the applications can understand the language in which the speech occurs.
- Classification of documents
  - Understanding of the natural language to identify the category of any document.
  - Large set of documents to be organized and arranged based upon the categories.

Thank You