7.2.4 Snowball stemming Algorithm:
It is a very useful stemming algorithm.NLTK has a **SnowballStemme**r class with the help of which we can implement Snowball Stemmer algorithms. It supports 15 languages other than English. For using this class, we have to build a sub-form with the language name that we are making use of and then use the stem() method on its successful creation.
Example: Bonjoura--->Bonjour

```
import nltk
from nltk.stem import SnowballStemmer
French_stemmer = SnowballStemmer('french')
French_stemmer.stem ('Bonjoura')
```

Output: 'Bonjour'

8. What is Lemmatization?

It is the process of assembling the inflected parts of a word such that they can be recognized as a single element, called the word's lemma or it's vocabulary form.[3] This process is the same as stemming but it adds meaning to particular words. In simple words, it connects text with alike meanings to a single word.
It is defined as an algorithm technique of finding the lemma of a word which is a root word rather than a root stem.[4] It is based on the intended meaning the word is trying to convey.
Example:
1. rocks: rock
2. Corpora: corpus
3. Better: good

8.1 Applications of Lemmatization:

- Use in Biomedicine: processing of text related to biomedicine can be efficient by using specialized lemmatization and may increase the efficiency of data retrieval tasks.
- Used in comprehensive retrieval systems like search engines
- Used in compact indexing

8.2 Implementation of lemmatization words using NLTK:

NLTK provides **WordNetLemmatizer** class which is a slim cover wrapped aroundthe **wordnetCorpus**. This class makes use of a function called M**orphy()** to the **WordNetCorpusReader** class to find a root word/lemma.

```
import nltk
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()
lemmatizer.lemmatize('books')
```

Output: 'book'

9. Difference between Lemmatization and Stemming

- On one hand, Stemming techniques are implemented by chopping off from either the end or beginning of the word-text, keeping track of commonly used prefixes and suffixes that could be available in an infected word-text, and On the other hand, lemmatization considers the study of the word-texts intending to find something which adds meaning to it.
- Lemma is the foundation of all its inflected parts, and a stem isn't. This is why orderly dictionaries are a record of lemmas and not of stems.
- examples in support

  a. **Stemming:**
  Form=studies, suffix = -es, stem = studi
  Form=studying, suffix = -ing, stem = study

  b. **Lemmatization:**
  Form=studies, Meaningful Information=Third person,singular number,present tense of the verb study, lemma=study
  Form=studying, Morphological Infomation=Gerund of verb study,Lemma=study

10. Which one is best: lemmatization or stemming?

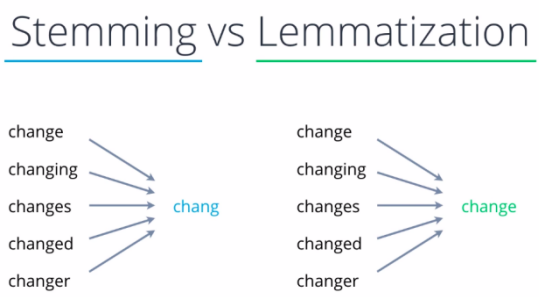

Figure 3: word stemming and Example of reduction in lemmatization.

Even though performing stemming is easier than the process lemmatization, the latter is proves to be the ultimate choice. Deep linguistics understanding is required to form the glossary that permits the algorithm to search for the meaningful part of the word in the process of lemmatization. Once this is done, the outcome will be more accurate. It makes use of vocabulary and morphological analysis of words to receive output free from derivative affixes.
On the other hand, stemming simply chops off the end of words irrespective of the fact that the output is conveyed any meaningful information.

CONCLUSION:

Processing of Natural Language is the area of study which focuses on making the communication between the language spoken by humans and machines possible. It helps to understand the text, enabling machines to recognize how a particular human talks in various languages concerning the place they reside. This human-like computer communication enables existent world implementations such as automatic text summary generation, the study of emotions, pulling out the relationship, and many more. Lemmatization and Stemming play a major role in text and natural language processing. Both of them create the base part of the inflected word-texts. The difference lies in the fact that stem is not a real word-text whereas lemma is a real language text format.

Acknowledgments

These should be brief and placed at the end of the text before the references.

REFERENCES

[1] Gelbukh, "Natural language processing," Fifth International Conference on Hybrid Intelligent Systems (HIS'05), Rio de Janeiro, Brazil, 2005, pp. 1 pp.-, DOI:10.1109/ICHIS.2005.79.
[2] Khurana, Diksha & Koli, Aditya & Khatter, Kiran & Singh, Sukhdev. (2017). Natural Language Processing: State of The Art, Current Trends and Challenges
[3] Balakrishnan, Vimala & Ethel, Lloyd-Yemoh. (2014). Stemming and Lemmatization: A Comparison of Retrieval Performances. Lecture Notes on Software Engineering. 2. 262-267. 10.7763/LNSE.2014.V2.134.
[4] Halácsy, P. (2006). Benefits of deep NLP-based Lemmatization for Information Retrieval. *CLEF*.
[5] Joseph, Sethunya & Sedimo, Kutlwano & Kaniwa, Freeson & Hlomani, Hlomani & Letsholo, Keletso. (2016). Natural Language Processing: A Review. Natural Language Processing: A Review. 6. 207-210.
[6] Sun, Shiliang & Luo, Chen & Chen, Junyu. (2016). A Review of Natural Language Processing Techniques for Opinion Mining Systems. Information Fusion. 36. 10.1016/j.inffus.2016.10.004.
[7] Krishna Prakash Kalyanathaya, D. Akila, and P. Rajesh. Advances in Natural Language Processing – A Survey of Current Research Trends, Development Tools, and Industry Applications. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5C, February 2019.
[8]https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html