Date:
Page:

## Lemmatization:-

i] In natural language processing, there may come a time when we want our program to recognize that the words 'ask' and 'asked' are just different tenses of same verb.

ii] This is the idea of reducing different forms of a word to a core root. words that are derived from one another can be mapped to a central word or symbol, especially if they have same core meaning.

iii] For grammatical reasons, documents are going to use different forms of a word, such as organize, organizes and organizing.

iv] Additionally, there are families of derivationally related words with similar meanings such as, democracy, democratic and democratization.

v] In many situations, it seems as if it would be useful for a search for one of these words to return documents that contain another word in the set.

vi] This is where like stemming or lemmatization comes in. The main goal of lemmatization is to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form.

vii] For instance:

am, are, is ⟶ be

car, cars, car's, cars' ⟶ car

viii] The result of this mapping of text will be something like:

the boy's cars are different colors ⇒

the boy car are differ color

ix] Lemmatization usually refers to doing things properly with the use of vocabulary and morphological analysis of words, normally aiming to remove the inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma.

x] Lemmatizations is a more calculated process. It involves resolving words to their dictionary form. In fact, a lemma of a word is its dictionary form.

xi] Because, lemmatization is more nuanced in this respect, it requires a little more to actually make work.

xii] For lemmatization to resolve a word to its lemma, it needs to know its parts-of-speech tagger. This allows is to do better resolutions (like resolving is and are to "be").

xiii] Another thing to note about lemmatization is that it's often times harder to create lemmatizer in a new language, than it is in Stemming algorithm

xiv] Because lemmatizer require a lot more knowledge about the structure of a language, it's a much more intensive process than just trying to set up a heuristic Stemming algorithm.

xv] If confronted with token saw, Stemming might return just S, whereas lemmatization would attempt to return either see or saw depending on whether the use of token was a verb or a noun.

xvi] lemmatizer, a tool from NLP which does full morphological analysis produces at most very modest benefits for retrival, and accurately identify lemma for each word.