Lematization

- to overcome the problem of the stemming we choose lematization
- firstly we tokenize the paragraph into diff senetnces then
- lematization is same kind of the stemming but intermediate represention/root form has a meaning
- ex : going,goes ,gone = go

there is a library in nltk for lematizing the text

```
In [1]: import nltk
        from nltk.stem import WordNetLemmatizer
        from nltk.corpus import stopwords
```

```
In [2]: paragraph=paragraph = """Thank you all so very much. Thank you to the Academy.
                        Thank you to all of you in this room. I have to congratulate
                        the other incredible nominees this year. The Revenant was
                        the product of the tireless efforts of an unbelievable cast
                        and crew. First off, to my brother in this endeavor, Mr. Tom
                        Hardy. Tom, your talent on screen can only be surpassed by
                        your friendship off screen … thank you for creating a t
                        ranscendent cinematic experience. Thank you to everybody at
                        Fox and New Regency … my entire team. I have to thank
                        everyone from the very onset of my career … To my parents;
                        none of this would be possible without you. And to my
                        friends, I love you dearly; you know who you are. And lastly,
                        I just want to say this: Making The Revenant was about
                        man's relationship to the natural world. A world that we
                        collectively felt in 2015 as the hottest year in recorded
                        history. Our production needed to move to the southern
                        tip of this planet just to be able to find snow. Climate
                        change is real, it is happening right now. It is the most
                        urgent threat facing our entire species, and we need to work
                        collectively together and stop procrastinating. We need to
                        support leaders around the world who do not speak for the
                        big polluters, but who speak for all of humanity, for the
                        indigenous people of the world, for the billions and
                        billions of underprivileged people out there who would be
                        most affected by this. For our children's children, and
                        for those people out there whose voices have been drowned
                        out by the politics of greed. I thank you all for this
                        amazing award tonight. Let us not take this planet for
                        granted. I do not take tonight for granted. Thank you so very much."""
```

convert the paragraphs into sentenses

```
In [4]: sentences = nltk.sent_tokenize(paragraph)
        lemmatizer = WordNetLemmatizer()
```

Lematization

```
In [6]: for i in range(len(sentences)):
            words = nltk.word_tokenize(sentences[i])
            words = [lemmatizer.lemmatize(word) for word in words if word not in set(stopwords.words
        ('english'))]
            sentences[i] = ' '.join(words)
```

```
In [7]: words
```

```
Out[7]: ['Thank', 'much', '.']
```

```
In [8]: sentences
```

```
Out[8]: ['Thank much .',
         'Thank Academy .',
         'Thank room .',
         'I congratulate incredible nominee year .',
         'The Revenant product tireless effort unbelievable cast crew .',
         'First , brother endeavor , Mr. Tom Hardy .',
         'Tom , talent screen surpassed friendship screen … thank creating ranscendent cinematic expe
        rience .',
         'Thank everybody Fox New Regency … entire team .',
         'I thank everyone onset career … To parent ; none would possible without .',
         'And friend , I love dearly ; know .',
         "And lastly , I want say : Making The Revenant man 's relationship natural world .",
         'A world collectively felt 2015 hottest year recorded history .',
         'Our production needed move southern tip planet able find snow .',
         'Climate change real , happening right .',
         'It urgent threat facing entire specie , need work collectively together stop procrastinatin
        g .',
         'We need support leader around world speak big polluter , speak humanity , indigenous people
        world , billion billion underprivileged people would affected .',
         'For child ' child , people whose voice drowned politics greed .',
         'I thank amazing award tonight .',
         'Let u take planet granted .',
         'I take tonight granted .',
         'Thank much .']
```

difference between lematization and stemming

lemmatization	Stemming
word representaions having meaning	word representaions having does not having meaning
takes ore time than stemming	takes less time
uses lemmatization when meaning of the words are important for analysis	uses stemming when meaning of the words are not important for analysis
ex:question and answer apps	ex: spam detection