

N-grams in NLTK

SECTION 3

Eng. Fatma

What Is N-Gram?

- They are basically a set of co-occurring words within a given window and when computing the n-grams you typically move one word forward (although you can move X words forward in more advanced scenarios).
- N can be 1, 2, or other positive integer.
- if N is 1, we call it a unigram;

Applications of N-Grams in NLP

- N-grams are widely used in various NLP tasks. Here are a few examples:
- **Language Modeling:** N-grams can be used to predict the next item in a sequence, making them useful for language models in speech recognition, typing prediction, and other generative tasks.
- **Text Classification:** They can serve as features for algorithms that classify documents into categories, such as spam filters or sentiment analysis.
- **Machine Translation:** N-grams help in statistical machine translation systems by providing probabilities of sequences of words appearing together.

Applications of N-Grams in NLP

- N-grams are widely used in various NLP tasks. Here are a few examples:
- **Spell Checking and Correction:** They can be used to suggest corrections for misspelled words based on the context provided by surrounding words.
- **Information Retrieval:** Search engines use n-grams to index texts and provide search results based on the likelihood of n-gram sequences.

Applications of N-Grams in NLP

- Speech recognition where phonemes and sequences of phonemes are modeled using a N-gram distribution.
- Parsing on words are modeled so that each N-gram is composed of N words. For language identification, sequences of characters/graphemes (e.g. letters of the alphabet) are modeled for different languages.
- Auto sentences completion
- Auto spell-check
- Semantic analysis

What Is N-Gram?

- **Example:**

- sentence “The cow jumps over the moon”. If **N=2** (known as **bigrams**), then the ngrams would be:
- the cow
- cow jumps
- jumps over
- over the
- the moon

What Is N-Gram?

- **Example:**

- If **N=3**, the n-grams would be:
- the cow jumps
- cow jumps over
- jumps over the
- over the moon.

How many N-grams in a sentence?

- If X =Num of words in a given sentence K , the number of n -grams for sentence K would be:

$$Ngrams_K = X - (N - 1)$$

Generation of N-Grams in NLTK

```
import nltk
from nltk import ngrams
sentence = input( "Enter the sentence: " )
n = int(input( "Enter the value of n: " ))
n_grams = ngrams(sentence.split(), n)
for grams in n_grams:
    print(grams)
```

Generation of N-Grams in NLTK

Enter the sentence: **To Sherlock Holmes she is always "The Woman". I have seldom heard him mention her under any other name.**

Enter the value of n: **2**

('To', 'Sherlock') ('Sherlock', 'Holmes') ('Holmes', 'she') ('she', 'is')
('is', 'always') ('always', '"The') ('"The', 'Woman".') ('Woman".', 'I')
('I', 'have') ('have', 'seldom') ('seldom', 'heard') ('heard', 'him')
('him', 'mention') ('mention', 'her') ('her', 'under') ('under', 'any')
('any', 'other') ('other', 'name.')

Generation of N-Grams in NLTK

Enter the sentence: **To Sherlock Holmes she is always "The Woman". I have seldom heard him mention her under any other name.**

Enter the value of n: **3**

('To', 'Sherlock', 'Holmes') ('Sherlock', 'Holmes', 'she')
('Holmes', 'she', 'is') ('she', 'is', 'always') ('is', 'always', '"The')
('always', '"The', 'Woman".') ('"The', 'Woman".', 'I')
('Woman".', 'I', 'have') ('I', 'have', 'seldom') ('have', 'seldom', 'heard')
('seldom', 'heard', 'him') ('heard', 'him', 'mention') ('him', 'mention', 'her')
('mention', 'her', 'under') ('her', 'under', 'any') ('under', 'any', 'other')
('any', 'other', 'name.')

Generation of N-Grams in NLTK

Enter the sentence: **To Sherlock Holmes she is always "The Woman". I have seldom heard him mention her under any other name.**

Enter the value of n: **4**

('To', 'Sherlock', 'Holmes', 'she') ('Sherlock', 'Holmes', 'she', 'is')
('Holmes', 'she', 'is', 'always') ('she', 'is', 'always', '"The')
('is', 'always', '"The', 'Woman".') ('always', '"The', 'Woman".', 'I')
('"The', 'Woman".', 'I', 'have') ('Woman".', 'I', 'have', 'seldom')
('I', 'have', 'seldom', 'heard') ('have', 'seldom', 'heard', 'him')
('seldom', 'heard', 'him', 'mention') ('heard', 'him', 'mention', 'her')
('him', 'mention', 'her', 'under') ('mention', 'her', 'under', 'any')
('her', 'under', 'any', 'other') ('under', 'any', 'other', 'name.')