# NLTK for english text :

import nltk, string

from nltk.tokenize import word\_tokenize

nltk.download('stopwords')

from nltk.corpus import stopwords

from nltk.stem.porter import PorterStemmer

from nltk import pos\_tag

nltk.download('averaged\_perceptron\_tagger')

nltk.download('punkt')

​

text = """Perhaps one of the most significant advances made by Arabic mathematics began at this time with the work of al-Khwarizmi, namely the beginnings of algebra. It is important to understand just how significant this new idea was. It was a revolutionary move away from the Greek concept of mathematics which was essentially geometry. Algebra was a unifying theory which allowed rational numbers, irrational numbers, geometrical magnitudes, etc., to all be treated as "algebraic objects". It gave mathematics a whole new development path so much broader in concept to that which had existed before, and provided a vehicle for future development of the subject. Another important aspect of the introduction of algebraic ideas was that it allowed mathematics to be applied to itself in a way which had not happened before."""

text = text.lower()

text\_p = "".join([char for char in text if char not in string.punctuation])

words = nltk.word\_tokenize(text\_p)

stop\_words = stopwords.words('english')

filtered\_words = [word for word in words if word not in stop\_words]

porter = PorterStemmer()

stemmed = [porter.stem(word) for word in filtered\_words]

pos = pos\_tag(filtered\_words)

for tree in pos:

print(tree)

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#NLTK for arabic text :

import nltk, string

from nltk.stem.isri import ISRIStemmer

from nltk.tokenize import word\_tokenize

nltk.download('stopwords')

from nltk.corpus import stopwords

from nltk.stem.porter import PorterStemmer

from nltk import pos\_tag

nltk.download('averaged\_perceptron\_tagger')

nltk.download('punkt')

nltk.download('maxent\_ne\_chunker')

nltk.download('words')

​

text = u'ربما كانت أحد أهم التطورات التي قامت بها الرياضيات العربية التي بدأت في هذا الوقت بعمل الخوارزمي و هي بدايات الجبر، و من المهم فهم كيف كانت هذه الفكرة الجديدة مهمة، فقد كانت خطوة ثورية بعيدا عن المفهوم اليوناني للرياضيات التي هي في جوهرها هندسة، الجبر كان نظرية موحدة تتيح الأعداد الكسرية و الأعداد اللا كسرية، و المقادير الهندسية و غيرها، أن تتعامل على أنها أجسام جبرية، و أعطت الرياضيات ككل مسارا جديدا للتطور بمفهوم أوسع بكثير من الذي كان موجودا من قبل، و قدم وسيلة للتنمية في هذا الموضوع مستقبلا و جانب آخر مهم لإدخال أفكار الجبر و هو أنه سمح بتطبيق الرياضيات على نفسها بطريقة لم تحدث من قبل'

text\_p = "".join([char for char in text if char not in string.punctuation])

"""isri\_stemmer = ISRIStemmer()

stem\_word = isri\_stemmer.stem(text)

print(stem\_word)"""

print(text\_p)

words = word\_tokenize(text\_p)

print(words)

arb\_stopwords = set(nltk.corpus.stopwords.words("arabic"))

filtered\_words = [word for word in words if word not in arb\_stopwords]

print(filtered\_words)

sentence = nltk.tokenize.sent\_tokenize(text)

# tokens = [nltk.tokenize.word\_tokenize(s) for s in sentence]

tokens = [nltk.tokenize.wordpunct\_tokenize(s) for s in sentence]

# Here pos tagging isn't right :'(

PosTokens = [nltk.pos\_tag(e) for e in tokens]

chunks = nltk.ne\_chunk\_sents(PosTokens)

for tree in chunks:

print(tree)

#porter = PorterStemmer()

#stemmed = [porter.stem(word) for word in filtered\_words]

#print(stemmed)

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Spacy for english

import spacy

from spacy.lang.en import English

nlp = spacy.load('en\_core\_web\_sm')

text = """Perhaps one of the most significant advances made by Arabic mathematics began at this time with the work of al-Khwarizmi, namely the beginnings of algebra. It is important to understand just how significant this new idea was. It was a revolutionary move away from the Greek concept of mathematics which was essentially geometry. Algebra was a unifying theory which allowed rational numbers, irrational numbers, geometrical magnitudes, etc., to all be treated as "algebraic objects". It gave mathematics a whole new development path so much broader in concept to that which had existed before, and provided a vehicle for future development of the subject. Another important aspect of the introduction of algebraic ideas was that it allowed mathematics to be applied to itself in a way which had not happened before."""

doc = nlp(text)

token\_list = []

for token in doc:

token\_list.append(token.text)

print (token\_list)

lemmas = [[token.text,token.lemma\_] for token in doc]

print(lemmas)

spacy\_stopwords = STOP\_WORDS

print(spacy\_stopwords)

filtered\_sent=[]

for word in doc:

if word.is\_stop==False and word.text.isalpha()==True:

filtered\_sent.append(word)

print(filtered\_sent)

post\_tag=[]

for post in doc:

post\_tag.append([post.text,post.pos\_])

print (post\_tag)

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Spacy for arabic text :

import spacy

from spacy.lang.ar import Arabic

nlp = Arabic()

text = u'ربما كانت أحد أهم التطورات التي قامت بها الرياضيات العربية التي بدأت في هذا الوقت بعمل الخوارزمي و هي بدايات الجبر، و من المهم فهم كيف كانت هذه الفكرة الجديدة مهمة، فقد كانت خطوة ثورية بعيدا عن المفهوم اليوناني للرياضيات التي هي في جوهرها هندسة، الجبر كان نظرية موحدة تتيح الأعداد الكسرية و الأعداد اللا كسرية، و المقادير الهندسية و غيرها، أن تتعامل على أنها أجسام جبرية، و أعطت الرياضيات ككل مسارا جديدا للتطور بمفهوم أوسع بكثير من الذي كان موجودا من قبل، و قدم وسيلة للتنمية في هذا الموضوع مستقبلا و جانب آخر مهم لإدخال أفكار الجبر و هو أنه سمح بتطبيق الرياضيات على نفسها بطريقة لم تحدث من قبل'

doc = nlp(text)

token\_list = []

for w in doc:

token\_list.append(w.text)

print (token\_list)

lemmas = [[token.text,token.lemma\_] for token in doc]

print(lemmas)

spacy\_stopwords = STOP\_WORDS

print(spacy\_stopwords)

filtered\_sent=[]

for word in doc:

if word.is\_stop==False and word.text.isalpha()==True:

filtered\_sent.append(word)

print(filtered\_sent)

post\_tag=[]

for post in doc:

post\_tag.append([post.text,post.pos\_])

print (post\_tag)