In [ ]:

```
## input text article
article_text="Just what is agility in the context of software engineering work? Ivar Jacobs
```

In [ ]:

```
import re
import nltk
from nltk.tokenize import word_tokenize
```

Type *Markdown* and LaTeX: α^2

In [ ]:

```
article_text = article_text.lower()
article_text
```

In [ ]:

```
# remove spaces, punctuations and numbers
clean_text = re.sub('[^a-zA-Z]', ' ', article_text)
clean_text = re.sub('\s+', ' ', clean_text)
clean_text
```

In [ ]:

```
# split into sentence list
sentence_list = nltk.sent_tokenize(article_text)
sentence_list
```

In [ ]:

```
tokens=word_tokenize(clean_text)
```

In [ ]:

```
tokens
```

In [ ]:

```
## run this cell once to download stopwords
# import nltk
#nltk.download('stopwords')
```

In [ ]:

```
stopwords = nltk.corpus.stopwords.words('english')
impword=[]
word_frequencies = {}
for word in tokens:
    if word not in stopwords:
        impword.append(word)
        if word not in word_frequencies:
            word_frequencies[word] = 1
        else:
            word_frequencies[word]+= 1
impword
```

In [ ]:

```
word_frequencies
```

In [ ]:

```
from nltk.util import ngrams
unigram=[]
bigram=[]
trigram=[]
```

In [ ]:

```
bigram.extend(list(ngrams(impword, 2,pad_left=True, pad_right=True)))
trigram.extend(list(ngrams(tokens, 3, pad_left=True, pad_right=True)))
unigram.extend(list(ngrams(tokens, 1, pad_left=True, pad_right=True)))
```

In [ ]:

```
trigram
```

In [ ]:

```
bigram
```

In [ ]:

```
unigram
```

In [ ]:

```

```