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
```

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 [ ]:

```
part_of_speech_tags=nltk.pos_tag(tokens)
part_of_speech_tags
```

In [ ]:

```
from sklearn.feature_extraction.text import CountVectorizer
```

In [ ]:

```
cv = CountVectorizer()
sents = ['coronavirus is a highly infectious disease',
   'coronavirus affects older people the most',
   'older people are at high risk due to this disease']
```

In [ ]:

```
X = cv.fit_transform(sents)
X = X.toarray()
X
```

In [ ]:

```
```