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 collections import Counter
```

In [ ]:

```
dictionary=Counter(word_freq)
```

In [ ]:

```
from nltk.stem import PorterStemmer
from nltk.stem import LancasterStemmer
```

In [ ]:

```
porter = PorterStemmer()
lancaster=LancasterStemmer()
#proide a word to be stemmed
print("Porter Stemmer")
print(porter.stem("cats"))
print(porter.stem("trouble"))
print(porter.stem("troubling"))
print(porter.stem("troubled"))
print("Lancaster Stemmer")
print(lancaster.stem("cats"))
print(lancaster.stem("trouble"))
print(lancaster.stem("troubling"))
print(lancaster.stem("troubled"))
```

In [ ]:

```
word_list = ["friend", "friendship", "friends", "friendships","stabil","destabilize","misun
print("{0:20}{1:20}{2:20}".format("Word","Porter Stemmer","lancaster Stemmer"))
for word in word_list:
    print("{0:20}{1:20}{2:20}".format(word,porter.stem(word),lancaster.stem(word)))
```

In [ ]:

```
stem_sentence=[]
for word in tokens:
    stem_sentence.append(porter.stem(word))
    stem_sentence.append(" ")
stem_sentence
```

In [ ]:

```
from nltk.stem import WordNetLemmatizer
lem1 = WordNetLemmatizer()
```

In [ ]:

```
print("{0:20}{1:20}".format("Word","Lemma"))
for word in impword:
    print ("{0:20}{1:20}".format(word,lem1.lemmatize(word)))
```