

### Step 1: Selection of Corpus

Select a corpus that could be either tweets or articles along with text excerpts. For simplicity:

- Chosen Corpus - "Alice's Adventures in Wonderland" (first 10 pages).

### Step 2: Text Cleaning

#### 1. Remove irrelevant content:

The text cleaning process involves removing headers and footers along with all metadata elements. For ease of use we are using the built-in functionality within Microsoft word to make the changes – it can easily replace characters and change all words to lowercase.

#### 2. Convert to lowercase:

o Example: "Hello World" → "hello world"

#### 3. Remove punctuations, numbers, and special characters:

o Example: "Hello, World! 123" → "hello world"

Example of basic cleaned Text: Alice pursued the white rabbit which led her into an unfamiliar wonderland

### Step 3: Tokenization

#### 1. Break text into individual words/tokens:

o Tokens: Alice pursued the white rabbit until she entered a bizarre wonderland.

#### 2. Count tokens:

o Total tokens: 1636

o Unique tokens: 1636

Reflection: The tokenization process serves as a critical step for dividing text into small segments that machines can process and comprehend. We used Python and Jupyter Notebook to Tokenize and print the tokenized words as well as output total word counts.

### Step 4: Stop Word Removal

#### 1. Examples of Common Stop Words include ['and', 'the', 'is', 'a', 'into'].

#### 2. Remove stop words:

o Tokens after removal: ['alice', 'followed', 'white', 'rabbit', 'strange', 'wonderland']

### 3. Count updated tokens:

- o Total tokens: 750

- o Unique tokens: 750

Reflection: Eliminating stop words directs attention to significant words which results in better processing efficiency.

### Step 5: Stemming/Lemmatization

#### 1. Choose Lemmatization – method `nltk.stem.WordNetLemmatizer`

- o Example: "followed" → "follow"

#### 2. Apply lemmatization:

Original Tokens: ['alice', 'beginning', 'get', 'tired', 'sitting', 'sister', 'bank', 'nothing', 'twice', 'peeped', 'book', 'sister', 'reading', 'pictures', 'conversations', 'use', 'book', 'thought', 'alice', 'without']

Lemmatized Tokens: ['alice', 'begin', 'get', 'tire', 'sit', 'sister', 'bank', 'nothing', 'twice', 'peep', 'book', 'sister', 'read', 'picture', 'conversations', 'use', 'book', 'think', 'alice', 'without']

Reflection: Lemmatization extracts the root meaning of words while maintaining their context which proves essential for sentence structure analysis.

### Step 6: Part-of-Speech (POS) Tagging

Manually tag 20 tokens:

1. **alice** (Proper Noun, NNP)
2. **begin** (Verb, VB)
3. **get** (Verb, VB)
4. **tire** (Verb, VB)
5. **sit** (Verb, VB)
6. **white** (Adjective, JJ)
7. **bank** (Noun, NN)
8. **nothing** (Pronoun, PRP)
9. **twice** (Adverb, RB)

10. **peep** (Verb, VB)
11. **book** (Noun, NN)
12. **sister** (Noun, NN)
13. **read** (Verb, VB)
14. **picture** (Noun, NN)
15. **conversations** (Noun, NNS)
16. **use** (Verb, VB)
17. **book** (Noun, NN)
18. **think** (Verb, VB)
19. **rabbit** (Noun, NN)
20. **without** (Preposition, IN)

Reflection: Through POS tagging we understand a word's function which allows interpretations that recognize context.

#### Step 7: Vectorization

##### 1. Choose TF-IDF:

- o Example: ['alice', 'follow', 'rabbit', 'wonderland']

- o TF-IDF values: The table displays TF-IDF values for each token representing their distribution across documents and their importance in a collection.

Reflection: The vectorization process converts text tokens into numerical forms enabling machine computation to understand human language.

Jupyter Notebook Commands and Output used for this assignment:

```
import pandas as pd
```

```
import nltk
```

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

```
from nltk.corpus import stopwords
```

In [4]:

```
csv_file = "AliceTokenization1Row.csv"
```

```
df = pd.read_csv(csv_file)
```

In [6]:

```
df = pd.read_csv(csv_file, header=None) # No headers in the file
```

```
df.columns = ['words'] # Add a column name for convenience
```

```
# Display the first few rows to confirm its structure
```

```
print(df.head())
```

```
words
```

```
0    alice
```

```
1     was
```

```
2  beginning
```

```
3      to
```

```
4     get
```

In [8]:

```
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
```

```
[nltk_data] C:\Users\Richard\AppData\Roaming\nltk_data...
```

```
[nltk_data] Package punkt is already up-to-date!
```

Out[8]:

```
True
```

In [10]:

```
tokens = df['words'].tolist()
```

In [12]:

```
print(tokens)
```

['alice', 'was', 'beginning', 'to', 'get', 'very', 'tired', 'of', 'sitting', 'by', 'her', 'sister', 'on', 'the', 'bank', 'and', 'of', 'having', 'nothing', 'to', 'do', 'once', 'or', 'twice', 'she', 'had', 'peeped', 'into', 'the', 'book', 'her', 'sister', 'was', 'reading', 'but', 'it', 'had', 'no', 'pictures', 'or', 'conversations', 'in', 'it', 'and', 'what', 'is', 'the', 'use', 'of', 'a', 'book', 'thought', 'alice', 'without', 'pictures', 'or', 'conversations', 'so', 'she', 'was', 'considering', 'in', 'her', 'own', 'mind', 'as', 'well', 'as', 'she', 'could', 'for', 'the', 'hot', 'day', 'made', 'her', 'feel', 'very', 'sleepy', 'and', 'stupid', 'whether', 'the', 'pleasure', 'of', 'making', 'a', 'daisychain', 'would', 'be', 'worth', 'the', 'trouble', 'of', 'getting', 'up', 'and', 'picking', 'the', 'daisies', 'when', 'suddenly', 'a', 'white', 'rabbit', 'with', 'pink', 'eyes', 'ran', 'close', 'by', 'her', 'there', 'was', 'nothing', 'so', 'very', 'remarkable', 'in', 'that', 'nor', 'did', 'alice', 'think', 'it', 'so', 'very', 'much', 'out', 'of', 'the', 'way', 'to', 'hear', 'the', 'rabbit', 'say', 'to', 'itself', 'oh', 'dear', 'oh', 'dear', 'i', 'shall', 'be', 'too', 'late', 'when', 'she', 'thought', 'it', 'over', 'afterwards', 'it', 'occurred', 'to', 'her', 'that', 'she', 'ought', 'to', 'have', 'wondered', 'at', 'this', 'but', 'at', 'the', 'time', 'it', 'all', 'seemed', 'quite', 'natural', 'but', 'when', 'the', 'rabbit', 'actually', 'took', 'a', 'watch', 'out', 'of', 'its', 'waistcoat', 'pocket', 'and', 'looked', 'at', 'it', 'and', 'then', 'hurried', 'on', 'alice', 'started', 'to', 'her', 'feet', 'for', 'it', 'flashed', 'across', 'her', 'mind', 'that', 'she', 'had', 'never', 'before', 'seen', 'a', 'rabbit', 'with', 'either', 'a', 'waistcoat', 'pocket', 'or', 'a', 'watch', 'to', 'take', 'out', 'of', 'it', 'and', 'burning', 'with', 'curiosity', 'she', 'ran', 'across', 'the', 'field', 'after', 'it', 'and', 'was', 'just', 'in', 'time', 'to', 'see', 'it', 'pop', 'down', 'a', 'large', 'rabbithole', 'under', 'the', 'hedge', 'in', 'another', 'moment', 'down', 'went', 'alice', 'after', 'it', 'never', 'once', 'considering', 'how', 'in', 'the', 'world', 'she', 'was', 'to', 'get', 'out', 'again', 'the', 'rabbithole', 'went', 'straight', 'on', 'like', 'a', 'tunnel', 'for', 'some', 'way', 'and', 'then', 'dipped', 'suddenly', 'down', 'so', 'suddenly', 'that', 'alice', 'had', 'not', 'a', 'moment', 'to', 'think', 'about', 'stopping', 'herself', 'before', 'she', 'found', 'herself', 'falling', 'down', 'what', 'seemed', 'to', 'be', 'a', 'very', 'deep', 'well', 'either', 'the', 'well', 'was', 'very', 'deep', 'or', 'she', 'fell', 'very', 'slowly', 'for', 'she', 'had', 'plenty', 'of', 'time', 'as', 'she', 'went', 'down', 'to', 'look', 'about', 'her', 'and', 'to', 'wonder', 'what', 'was', 'going', 'to', 'happen', 'next', 'first', 'she', 'tried', 'to', 'look', 'down', 'and', 'make', 'out', 'what', 'she', 'was', 'coming', 'to', 'but', 'it', 'was', 'too', 'dark', 'to', 'see', 'anything', 'then', 'she', 'looked', 'at', 'the', 'sides', 'of', 'the', 'well', 'and', 'noticed', 'that', 'they', 'were', 'filled', 'with', 'cupboards', 'and', 'bookshelves', 'here', 'and', 'there', 'she', 'saw', 'maps', 'and', 'pictures', 'hung', 'upon', 'pegs', 'she', 'took', 'down', 'a', 'jar', 'from', 'one', 'of', 'the', 'shelves', 'as', 'she', 'passed', 'it', 'was', 'labelled', 'orange', 'marmalade', 'but', 'to', 'her', 'great', 'disappointment', 'it', 'was', 'empty', 'she', 'did', 'not', 'like', 'to', 'drop', 'the', 'jar', 'for', 'fear', 'of', 'killing', 'somebody', 'underneath', 'so', 'managed', 'to', 'put', 'it', 'into', 'one', 'of', 'the', 'cupboards', 'as', 'she', 'fell', 'past', 'it', 'well', 'thought', 'alice', 'to', 'herself', 'after', 'such', 'a', 'fall', 'as', 'this', 'i', 'shall', 'think', 'nothing', 'of', 'tumbling', 'down', 'stairs', 'how', 'brave', 'theyll', 'all', 'think', 'me', 'at', 'home', 'why', 'i', 'wouldnt', 'say', 'anything', 'about', 'it', 'even', 'if', 'i', 'fell', 'off', 'the', 'top', 'of', 'the', 'house', 'which', 'was', 'very', 'likely', 'TRUE', 'down', 'down', 'down', 'would', 'the', 'fall', 'never', 'come', 'to', 'an', 'end', 'i', 'wonder', 'how', 'many', 'miles', 'i', 've', 'fallen', 'by', 'this', 'time', 'she', 'said',

'aloud', 'i', 'must', 'be', 'getting', 'somewhere', 'near', 'the', 'centre', 'of', 'the', 'earth', 'let', 'me', 'see', 'that', 'would', 'be', 'four', 'thousand', 'miles', 'down', 'i', 'think', 'for', 'you', 'see', 'alice', 'had', 'learnt', 'several', 'things', 'of', 'this', 'sort', 'in', 'her', 'lessons', 'in', 'the', 'schoolroom', 'and', 'though', 'this', 'was', 'not', 'a', 'very', 'good', 'opportunity', 'for', 'showing', 'off', 'her', 'knowledge', 'as', 'there', 'was', 'no', 'one', 'to', 'listen', 'to', 'her', 'still', 'it', 'was', 'good', 'practice', 'to', 'say', 'it', 'over', 'yes', 'thats', 'about', 'the', 'right', 'distancebut', 'then', 'i', 'wonder', 'what', 'latitude', 'or', 'longitude', 'ive', 'got', 'to', 'alice', 'had', 'not', 'the', 'slightest', 'idea', 'what', 'latitude', 'was', 'or', 'longitude', 'either', 'but', 'she', 'thought', 'they', 'were', 'nice', 'grand', 'words', 'to', 'say', 'presently', 'she', 'began', 'again', 'i', 'wonder', 'if', 'i', 'shall', 'fall', 'right', 'through', 'the', 'earth', 'how', 'funny', 'itll', 'seem', 'to', 'come', 'out', 'among', 'the', 'people', 'that', 'walk', 'with', 'their', 'heads', 'downwards', 'the', 'antipathies', 'i', 'think', 'she', 'was', 'rather', 'glad', 'there', 'was', 'no', 'one', 'listening', 'this', 'time', 'as', 'it', 'didnt', 'sound', 'at', 'all', 'the', 'right', 'word', 'but', 'i', 'shall', 'have', 'to', 'ask', 'them', 'what', 'the', 'name', 'of', 'the', 'country', 'is', 'you', 'know', 'please', 'maam', 'is', 'this', 'new', 'zealand', 'or', 'australia', 'and', 'she', 'tried', 'to', 'curtsey', 'as', 'she', 'spokefancy', 'curtseying', 'as', 'youre', 'falling', 'through', 'the', 'air', 'do', 'you', 'think', 'you', 'could', 'manage', 'it', 'and', 'what', 'an', 'ignorant', 'little', 'girl', 'shell', 'think', 'me', 'for', 'asking', 'no', 'itll', 'never', 'do', 'to', 'ask', 'perhaps', 'i', 'shall', 'see', 'it', 'written', 'up', 'somewhere', 'down', 'down', 'down', 'there', 'was', 'nothing', 'else', 'to', 'do', 'so', 'alice', 'soon', 'began', 'talking', 'again', 'dinah', 'll', 'miss', 'me', 'very', 'much', 'tonight', 'i', 'should', 'think', 'dinah', 'was', 'the', 'cat', 'i', 'hope', 'theyll', 'remember', 'her', 'saucer', 'of', 'milk', 'at', 'teatime', 'dinah', 'my', 'dear', 'i', 'wish', 'you', 'were', 'down', 'here', 'with', 'me', 'there', 'are', 'no', 'mice', 'in', 'the', 'air', 'im', 'afraid', 'but', 'you', 'might', 'catch', 'a', 'bat', 'and', 'thats', 'very', 'like', 'a', 'mouse', 'you', 'know', 'but', 'do', 'cats', 'eat', 'bats', 'i', 'wonder', 'and', 'here', 'alice', 'began', 'to', 'get', 'rather', 'sleepy', 'and', 'went', 'on', 'saying', 'to', 'herself', 'in', 'a', 'dreamy', 'sort', 'of', 'way', 'do', 'cats', 'eat', 'bats', 'do', 'cats', 'eat', 'bats', 'and', 'sometimes', 'do', 'bats', 'eat', 'cats', 'for', 'you', 'see', 'as', 'she', 'couldnt', 'answer', 'either', 'question', 'it', 'didnt', 'much', 'matter', 'which', 'way', 'she', 'put', 'it', 'she', 'felt', 'that', 'she', 'was', 'dozing', 'off', 'and', 'had', 'just', 'begun', 'to', 'dream', 'that', 'she', 'was', 'walking', 'hand', 'in', 'hand', 'with', 'dinah', 'and', 'was', 'saying', 'to', 'her', 'very', 'earnestly', 'now', 'dinah', 'tell', 'me', 'the', 'truth', 'did', 'you', 'ever', 'eat', 'a', 'bat', 'when', 'suddenly', 'thump', 'thump', 'down', 'she', 'came', 'upon', 'a', 'heap', 'of', 'sticks', 'and', 'dry', 'leaves', 'and', 'the', 'fall', 'was', 'over', 'navigate', 'full', 'screen', 'on/off', 'control', 'close', 'book', 'rabbithole', 'alice', 'was', 'not', 'a', 'bit', 'hurt', 'and', 'she', 'jumped', 'up', 'on', 'to', 'her', 'feet', 'in', 'a', 'moment', 'she', 'looked', 'up', 'but', 'it', 'was', 'all', 'dark', 'overhead', 'before', 'her', 'was', 'another', 'long', 'passage', 'and', 'the', 'white', 'rabbit', 'was', 'still', 'in', 'sight', 'hurrying', 'down', 'it', 'there', 'was', 'not', 'a', 'moment', 'to', 'be', 'lost', 'away', 'went', 'alice', 'like', 'the', 'wind', 'and', 'was', 'just', 'in', 'time', 'to', 'hear', 'it', 'say', 'as', 'it', 'turned', 'a', 'corner', 'oh', 'my', 'ears', 'and', 'whiskers', 'how', 'late', 'its', 'getting', 'she', 'was', 'close', 'behind', 'it', 'when', 'she', 'turned', 'the', 'corner', 'but', 'the', 'rabbit', 'was', 'no', 'longer',

'to', 'be', 'seen', 'she', 'found', 'herself', 'in', 'a', 'long', 'low', 'hall', 'which', 'was', 'lit', 'up', 'by', 'a', 'row', 'of', 'lamps', 'hanging', 'from', 'the', 'roof', 'there', 'were', 'doors', 'all', 'round', 'the', 'hall', 'but', 'they', 'were', 'all', 'locked', 'and', 'when', 'alice', 'had', 'been', 'all', 'the', 'way', 'down', 'one', 'side', 'and', 'up', 'the', 'other', 'trying', 'every', 'door', 'she', 'walked', 'sadly', 'down', 'the', 'middle', 'wondering', 'how', 'she', 'was', 'ever', 'to', 'get', 'out', 'again', 'suddenly', 'she', 'came', 'upon', 'a', 'little', 'threelegged', 'table', 'all', 'made', 'of', 'solid', 'glass', 'there', 'was', 'nothing', 'on', 'it', 'but', 'a', 'tiny', 'golden', 'key', 'and', 'alices', 'first', 'idea', 'was', 'that', 'this', 'might', 'belong', 'to', 'one', 'of', 'doors', 'of', 'the', 'hall', 'but', 'alas', 'either', 'the', 'locks', 'were', 'too', 'large', 'or', 'the', 'key', 'was', 'too', 'small', 'but', 'at', 'any', 'rate', 'it', 'would', 'not', 'open', 'any', 'of', 'them', 'however', 'on', 'the', 'second', 'time', 'round', 'she', 'came', 'upon', 'a', 'low', 'curtain', 'she', 'had', 'not', 'noticed', 'before', 'and', 'behind', 'it', 'was', 'a', 'little', 'door', 'about', 'fifteen', 'inches', 'high', 'she', 'tried', 'the', 'little', 'golden', 'key', 'in', 'the', 'lock', 'and', 'to', 'her', 'great', 'delight', 'it', 'fitted', 'alice', 'opened', 'the', 'door', 'and', 'found', 'that', 'it', 'led', 'into', 'a', 'small', 'passage', 'not', 'much', 'larger', 'than', 'a', 'rathole', 'she', 'knelt', 'down', 'and', 'looked', 'along', 'the', 'passage', 'into', 'the', 'loveliest', 'garden', 'you', 'ever', 'saw', 'how', 'she', 'longed', 'to', 'get', 'out', 'of', 'that', 'dark', 'hall', 'and', 'wander', 'about', 'among', 'those', 'beds', 'of', 'bright', 'flowers', 'and', 'those', 'cool', 'fountains', 'but', 'she', 'could', 'not', 'even', 'get', 'her', 'head', 'though', 'the', 'doorway', 'and', 'even', 'if', 'my', 'head', 'would', 'go', 'through', 'thought', 'poor', 'alice', 'it', 'would', 'be', 'of', 'very', 'little', 'use', 'without', 'my', 'shoulders', 'oh', 'how', 'i', 'wish', 'i', 'could', 'shut', 'up', 'like', 'a', 'telescope', 'i', 'think', 'i', 'could', 'if', 'i', 'only', 'knew', 'how', 'to', 'begin', 'for', 'you', 'see', 'so', 'many', 'outoftheway', 'things', 'had', 'happened', 'lately', 'that', 'alice', 'had', 'begun', 'to', 'think', 'that', 'very', 'few', 'things', 'indeed', 'were', 'really', 'impossible', 'there', 'seemed', 'to', 'be', 'no', 'use', 'in', 'waiting', 'by', 'the', 'little', 'door', 'so', 'she', 'went', 'back', 'to', 'the', 'table', 'half', 'hoping', 'she', 'might', 'find', 'another', 'key', 'on', 'it', 'or', 'at', 'any', 'rate', 'a', 'book', 'of', 'rules', 'for', 'shutting', 'people', 'up', 'like', 'telescopes', 'this', 'time', 'she', 'found', 'a', 'little', 'bottle', 'on', 'it', 'which', 'certainly', 'was', 'not', 'here', 'before', 'said', 'alice', 'and', 'tied', 'round', 'the', 'neck', 'of', 'the', 'bottle', 'was', 'a', 'paper', 'label', 'with', 'the', 'words', 'drink', 'me', 'beautifully', 'printed', 'on', 'it', 'in', 'large', 'letters', 'it', 'was', 'all', 'very', 'well', 'to', 'say', 'drink', 'me', 'but', 'the', 'wise', 'little', 'alice', 'was', 'not', 'going', 'to', 'do', 'that', 'in', 'a', 'hurry', 'no', 'ill', 'look', 'first', 'she', 'said', 'and', 'see', 'whether', 'its', 'marked', 'poison', 'or', 'not', 'for', 'she', 'had', 'read', 'several', 'nice', 'little', 'stories', 'about', 'children', 'who', 'had', 'got', 'burnt', 'and', 'eaten', 'up', 'by', 'wild', 'beasts', 'and', 'other', 'unpleasant', 'things', 'all', 'because', 'they', 'would', 'not', 'remember', 'the', 'simple', 'rules', 'their', 'friends', 'had', 'taught', 'them', 'such', 'as', 'that', 'a', 'redhot', 'poker', 'will', 'burn', 'you', 'if', 'you', 'hold', 'it', 'too', 'long', 'and', 'that', 'if', 'you', 'cut', 'your', 'finger', 'very', 'deeply', 'with', 'a', 'knife', 'it', 'usually', 'bleeds', 'and', 'she', 'had', 'never', 'forgotten', 'that', 'if', 'you', 'drink', 'much', 'from', 'a', 'bottle', 'marked', 'poison', 'it', 'is', 'almost', 'certain', 'to', 'disagree', 'with', 'you', 'sooner', 'or', 'later', 'however', 'this', 'bottle', 'was', 'not', 'marked', 'poison', 'so', 'alice', 'ventured', 'to',

'taste', 'it', 'and', 'finding', 'it', 'very', 'nice', 'it', 'had', 'in', 'fact', 'a', 'sort', 'of', 'mixed', 'flavour', 'of', 'cherrytart', 'custard', 'pineapple', 'roast', 'turkey', 'toffee', 'and', 'hot', 'battered', 'toast', 'she', 'very', 'soon', 'finished', 'it', 'off']

In [14]:

```
nltk.download('stopwords')
```

```
stop_words = set(stopwords.words('english'))
```

```
filtered_tokens = [word for word in tokens if word not in stop_words]
```

[nltk\_data] Downloading package stopwords to

[nltk\_data] C:\Users\Richard\AppData\Roaming\nltk\_data...

[nltk\_data] Package stopwords is already up-to-date!

In [16]:

```
print(filtered_tokens)
```

['alice', 'beginning', 'get', 'tired', 'sitting', 'sister', 'bank', 'nothing', 'twice', 'peeped', 'book', 'sister', 'reading', 'pictures', 'conversations', 'use', 'book', 'thought', 'alice', 'without', 'pictures', 'conversations', 'considering', 'mind', 'well', 'could', 'hot', 'day', 'made', 'feel', 'sleepy', 'stupid', 'whether', 'pleasure', 'making', 'daisychain', 'would', 'worth', 'trouble', 'getting', 'picking', 'daisies', 'suddenly', 'white', 'rabbit', 'pink', 'eyes', 'ran', 'close', 'nothing', 'remarkable', 'alice', 'think', 'much', 'way', 'hear', 'rabbit', 'say', 'oh', 'dear', 'oh', 'dear', 'shall', 'late', 'thought', 'afterwards', 'occurred', 'ought', 'wondered', 'time', 'seemed', 'quite', 'natural', 'rabbit', 'actually', 'took', 'watch', 'waistcoat', 'pocket', 'looked', 'hurried', 'alice', 'started', 'feet', 'flashed', 'across', 'mind', 'never', 'seen', 'rabbit', 'either', 'waistcoat', 'pocket', 'watch', 'take', 'burning', 'curiosity', 'ran', 'across', 'field', 'time', 'see', 'pop', 'large', 'rabbithole', 'hedge', 'another', 'moment', 'went', 'alice', 'never', 'considering', 'world', 'get', 'rabbithole', 'went', 'straight', 'like', 'tunnel', 'way', 'dipped', 'suddenly', 'suddenly', 'alice', 'moment', 'think', 'stopping', 'found', 'falling', 'seemed', 'deep', 'well', 'either', 'well', 'deep', 'fell', 'slowly', 'plenty', 'time', 'went', 'look', 'wonder', 'going', 'happen', 'next', 'first', 'tried', 'look', 'make', 'coming', 'dark', 'see', 'anything', 'looked', 'sides', 'well', 'noticed', 'filled', 'cupboards', 'bookshelves', 'saw', 'maps', 'pictures', 'hung', 'upon', 'pegs', 'took', 'jar', 'one', 'shelves', 'passed', 'labelled', 'orange', 'marmalade', 'great', 'disappointment', 'empty', 'like', 'drop', 'jar', 'fear', 'killing', 'somebody', 'underneath', 'managed', 'put', 'one', 'cupboards', 'fell', 'past', 'well', 'thought', 'alice', 'fall', 'shall', 'think', 'nothing', 'tumbling', 'stairs', 'brave', 'theyll', 'think', 'home', 'wouldnt', 'say', 'anything', 'even', 'fell', 'top', 'house', 'likely', 'TRUE', 'would', 'fall', 'never', 'come', 'end', 'wonder', 'many', 'miles', 'fallen', 'time', 'said', 'aloud', 'must', 'getting',



'somewhere', 'near', 'centre', 'earth', 'let', 'see', 'would', 'four', 'thousand', 'miles', 'think', 'see', 'alice', 'learnt', 'several', 'things', 'sort', 'lessons', 'schoolroom', 'though', 'good', 'opportunity', 'showing', 'knowledge', 'one', 'listen', 'still', 'good', 'practice', 'say', 'yes', 'thats', 'right', 'distancebut', 'wonder', 'latitude', 'longitude', 'ive', 'got', 'alice', 'slightest', 'idea', 'latitude', 'longitude', 'either', 'thought', 'nice', 'grand', 'words', 'say', 'presently', 'began', 'wonder', 'shall', 'fall', 'right', 'earth', 'funny', 'itll', 'seem', 'come', 'among', 'people', 'walk', 'heads', 'downwards', 'antipathies', 'think', 'rather', 'glad', 'one', 'listening', 'time', 'didnt', 'sound', 'right', 'word', 'shall', 'ask', 'name', 'country', 'know', 'please', 'maam', 'new', 'zealand', 'australia', 'tried', 'curtsey', 'spokefancy', 'curtseying', 'youre', 'falling', 'air', 'think', 'could', 'manage', 'ignorant', 'little', 'girl', 'shell', 'think', 'asking', 'itll', 'never', 'ask', 'perhaps', 'shall', 'see', 'written', 'somewhere', 'nothing', 'else', 'alice', 'soon', 'began', 'talking', 'dinah', 'miss', 'much', 'tonight', 'think', 'dinah', 'cat', 'hope', 'theyll', 'remember', 'saucer', 'milk', 'teatime', 'dinah', 'dear', 'wish', 'mice', 'air', 'im', 'afraid', 'might', 'catch', 'bat', 'thats', 'like', 'mouse', 'know', 'cats', 'eat', 'bats', 'wonder', 'alice', 'began', 'get', 'rather', 'sleepy', 'went', 'saying', 'dreamy', 'sort', 'way', 'cats', 'eat', 'bats', 'cats', 'eat', 'bats', 'sometimes', 'bats', 'eat', 'cats', 'see', 'couldnt', 'answer', 'either', 'question', 'didnt', 'much', 'matter', 'way', 'put', 'felt', 'dozing', 'begun', 'dream', 'walking', 'hand', 'hand', 'dinah', 'saying', 'earnestly', 'dinah', 'tell', 'truth', 'ever', 'eat', 'bat', 'suddenly', 'thump', 'thump', 'came', 'upon', 'heap', 'sticks', 'dry', 'leaves', 'fall', 'navigate', 'full', 'screen', 'on/off', 'control', 'close', 'book', 'rabbithole', 'alice', 'bit', 'hurt', 'jumped', 'feet', 'moment', 'looked', 'dark', 'overhead', 'another', 'long', 'passage', 'white', 'rabbit', 'still', 'sight', 'hurrying', 'moment', 'lost', 'away', 'went', 'alice', 'like', 'wind', 'time', 'hear', 'say', 'turned', 'corner', 'oh', 'ears', 'whiskers', 'late', 'getting', 'close', 'behind', 'turned', 'corner', 'rabbit', 'longer', 'seen', 'found', 'long', 'low', 'hall', 'lit', 'row', 'lamps', 'hanging', 'roof', 'doors', 'round', 'hall', 'locked', 'alice', 'way', 'one', 'side', 'trying', 'every', 'door', 'walked', 'sadly', 'middle', 'wondering', 'ever', 'get', 'suddenly', 'came', 'upon', 'little', 'threelegged', 'table', 'made', 'solid', 'glass', 'nothing', 'tiny', 'golden', 'key', 'alices', 'first', 'idea', 'might', 'belong', 'one', 'doors', 'hall', 'alas', 'either', 'locks', 'large', 'key', 'small', 'rate', 'would', 'open', 'however', 'second', 'time', 'round', 'came', 'upon', 'low', 'curtain', 'noticed', 'behind', 'little', 'door', 'fifteen', 'inches', 'high', 'tried', 'little', 'golden', 'key', 'lock', 'great', 'delight', 'fitted', 'alice', 'opened', 'door', 'found', 'led', 'small', 'passage', 'much', 'larger', 'rathole', 'knelt', 'looked', 'along', 'passage', 'loveliest', 'garden', 'ever', 'saw', 'longed', 'get', 'dark', 'hall', 'wander', 'among', 'beds', 'bright', 'flowers', 'cool', 'fountains', 'could', 'even', 'get', 'head', 'though', 'doorway', 'even', 'head', 'would', 'go', 'thought', 'poor', 'alice', 'would', 'little', 'use', 'without', 'shoulders', 'oh', 'wish', 'could', 'shut', 'like', 'telescope', 'think', 'could', 'knew', 'begin', 'see', 'many', 'outoftheway', 'things', 'happened', 'lately', 'alice', 'begun', 'think', 'things', 'indeed', 'really', 'impossible', 'seemed', 'use', 'waiting', 'little', 'door', 'went', 'back', 'table', 'half', 'hoping', 'might', 'find', 'another', 'key', 'rate', 'book', 'rules', 'shutting', 'people', 'like', 'telescopes', 'time', 'found', 'little', 'bottle', 'certainly', 'said', 'alice', 'tied', 'round', 'neck', 'bottle', 'paper', 'label', 'words', 'drink', 'beautifully', 'printed', 'large', 'letters', 'well', 'say',

'drink', 'wise', 'little', 'alice', 'going', 'hurry', 'ill', 'look', 'first', 'said', 'see', 'whether', 'marked', 'poison', 'read', 'several', 'nice', 'little', 'stories', 'children', 'got', 'burnt', 'eaten', 'wild', 'beasts', 'unpleasant', 'things', 'would', 'remember', 'simple', 'rules', 'friends', 'taught', 'redhot', 'poker', 'burn', 'hold', 'long', 'cut', 'finger', 'deeply', 'knife', 'usually', 'bleeds', 'never', 'forgotten', 'drink', 'much', 'bottle', 'marked', 'poison', 'almost', 'certain', 'disagree', 'sooner', 'later', 'however', 'bottle', 'marked', 'poison', 'alice', 'ventured', 'taste', 'finding', 'nice', 'fact', 'sort', 'mixed', 'flavour', 'cherrytart', 'custard', 'pineapple', 'roast', 'turkey', 'toffy', 'hot', 'battered', 'toast', 'soon', 'finished']

In [20]:

```
token_word_count = len(filtered_tokens)
```

```
print(f"Total Tokenized Word Count: {token_word_count}")
```

Total Tokenized Word Count: 750

In [22]:

```
total_count_before_stop = len(tokens)
```

```
print(f"Total Tokenized Word Count before stop words removed:  
{total_count_before_stop}")
```

Total Tokenized Word Count before stop words removed: 1636

In [32]:

```
from nltk.stem import WordNetLemmatizer
```

```
from nltk.corpus import wordnet
```

```
nltk.download('wordnet')
```

```
nltk.download('omw-1.4')
```

```
[nltk_data] Downloading package wordnet to
```

```
[nltk_data] C:\Users\Richard\AppData\Roaming\nltk_data...
```

```
[nltk_data] Package wordnet is already up-to-date!
```

```
[nltk_data] Downloading package omw-1.4 to
```

```
[nltk_data] C:\Users\Richard\AppData\Roaming\nltk_data...
```

[nltk\_data] Package omw-1.4 is already up-to-date!

Out[32]:

True

In [34]:

```
lemmatizer = WordNetLemmatizer()
```

In [36]:

```
lemmatized_tokens = [lemmatizer.lemmatize(word, pos=wordnet.VERB) for word in  
filtered_tokens]
```

In [38]:

```
print("Original Tokens:", filtered_tokens[:20])
```

```
Original Tokens: ['alice', 'beginning', 'get', 'tired', 'sitting', 'sister', 'bank', 'nothing', 'twice',  
'peeped', 'book', 'sister', 'reading', 'pictures', 'conversations', 'use', 'book', 'thought', 'alice',  
'without']
```

In [40]:

```
print("Lemmatized Tokens:", lemmatized_tokens[:20])
```

```
Lemmatized Tokens: ['alice', 'begin', 'get', 'tire', 'sit', 'sister', 'bank', 'nothing', 'twice', 'peep',  
'book', 'sister', 'read', 'picture', 'conversations', 'use', 'book', 'think', 'alice', 'without']
```

In [44]:

```
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
tfidf_vectorizer = TfidfVectorizer()
```

In [46]:

```
sample_text = ['alice', 'follow', 'rabbit', 'wonderland']
```

```
tfidf_output = tfidf_vectorizer.fit_transform(sample_text)
```

In [48]:

```
print("Feature Names:", tfidf_vectorizer.get_feature_names_out())
```

```
Feature Names: ['alice' 'follow' 'rabbit' 'wonderland']
```

In [50]:

ITAI-1370

Richard Rodriguez, Spencer Van Horne

Team: Automation Learners

```
print("TF-IDF Representation:\n", tfidf_output.toarray())
```

TF-IDF Representation:

```
[[1. 0. 0. 0.]
```

```
[0. 1. 0. 0.]
```

```
[0. 0. 1. 0.]
```

```
[0. 0. 0. 1.]]
```