In [1]:
```
import re
import nltk
from nltk.tokenize import word_tokenize

nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\ASUS\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

Out[1]:
```
True
```

In [7]:
```
text = """
Hello, you can reach me at john.doe@example.com and my colleague at java_smith
Another email is test.email @domain.net, and support can be contacted at help@
"""

email_pattern = "[a-zA-Z0-9._%+-]+@[a-zA-Z]+\.[a-zA-Z]{2,}"

emails = re.findall(email_pattern, text)

username = [email.split("@")[0] for email in emails]

print(f"Extracted emails: {emails}\nExtracted usernames: {username}")
```

```
Extracted emails: ['john.doe@example.com', 'java_smith123@company.org', 'help
@service.com']
Extracted usernames: ['john.doe', 'java_smith123', 'help']
```

Extracting top common words- extract the top 10 most common words in the text excluding stopwords

In [8]:
```
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from collections import Counter

nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\ASUS\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[8]:
```
True
```

In [14]:
```
text = """
Kingdom is a Japanese manga series written and illustrated by Yasuhisa Hara. I
since 2006 and has over 70 volumes. The story is set in the Warring States per
, a war orphan who dreams of becoming the greatest general under heaven. The s
, who aspires to rise above his lowly status and achieve greatness as a genera
(Ying Zheng), the young king of Qin, who aims to unify China. Together, they e
political intrigue, and personal growth.
"""

tokens = word_tokenize(text)

words = [token.lower() for token in tokens]

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

filtered_words = [word for word in words if word not in stop_words]

word_counts = Counter(filtered_words)

top_10_common = word_counts.most_common(10)

print("Top 10 most common words: ")
for word, count in top_10_common:
    print(f"{word}: {count}")
```

```
Top 10 most common words:
,: 8
.: 6
young: 2
story: 2
china: 2
shin: 2
(: 2
): 2
war: 2
orphan: 2
```

In [ ]:
```

```