

```
In [ ]: # Once, download the gutenber corpus and import libraries
import nltk
nltk.download('gutenberg')

# Set the correct path to the java.exe file
import os
java_path = "C:\Program Files\Common Files\Oracle\Java\javapath\java.exe"
os.environ['JAVAHOME'] = java_path
```

```
<>:9: SyntaxWarning: invalid escape sequence '\P'
<>:9: SyntaxWarning: invalid escape sequence '\P'
C:\Users\stefa\AppData\Local\Temp\ipykernel_19844\1952149573.py:9: SyntaxWarning: invalid escape sequence '\P'
  java_path = "C:\Program Files\Common Files\Oracle\Java\javapath\java.exe"
[nltk_data] Downloading package gutenber to
[nltk_data]   C:\Users\stefa\AppData\Roaming\nltk_data...
[nltk_data]   Package gutenber is already up-to-date!
```

```
In [25]: # Get all ids of the webtext corpus
ids = nltk.corpus.gutenberg.fileids()

lines = nltk.corpus.gutenberg.raw("shakespeare-hamlet.txt")

# Save webtext to a file, to make it easier to work with
with open("hamlet.txt", "w", encoding="utf-8") as f:
    f.write(lines)
```

```
In [32]: # Use 4-classes tagger NER to extract named entities
from nltk.tag import StanfordNERTagger
classifier_3_path = "C:\VScode Git\DS2\Assignment II\stanford-ner-2020-11-17\classifiers\english.all.3class.distsim.crf.ser.gz"
classifier_4_path = "C:\VScode Git\DS2\Assignment II\stanford-ner-2020-11-17\classifiers\english.conll.4class.distsim.crf.ser.gz"
classifier_7_path = "C:\VScode Git\DS2\Assignment II\stanford-ner-2020-11-17\classifiers\english.muc.7class.distsim.crf.ser.gz"
jar_path = "C:\VScode Git\DS2\Assignment II\stanford-ner-2020-11-17\stanford-ner.jar"

class_3 = StanfordNERTagger(classifier_3_path, jar_path).tag(lines.split())
class_4 = StanfordNERTagger(classifier_4_path, jar_path).tag(lines.split())
class_7 = StanfordNERTagger(classifier_7_path, jar_path).tag(lines.split())

filtered_class_3 = [x[0] for x in class_3 if x[1] == 'PERSON']
filtered_class_4 = [x[0] for x in class_4 if x[1] == 'PERSON']
filtered_class_7 = [x[0] for x in class_7 if x[1] == 'PERSON']

print("3-class tagger:")
print(filtered_class_3)
print(len(filtered_class_3))

print("4-class tagger:")
print(filtered_class_4)
print(len(filtered_class_4))

print("7-class tagger:")
print(filtered_class_7)
print(len(filtered_class_7))
```

```
<>:3: SyntaxWarning: invalid escape sequence '\V'
<>:4: SyntaxWarning: invalid escape sequence '\V'
<>:5: SyntaxWarning: invalid escape sequence '\V'
<>:6: SyntaxWarning: invalid escape sequence '\V'
<>:3: SyntaxWarning: invalid escape sequence '\V'
<>:4: SyntaxWarning: invalid escape sequence '\V'
<>:5: SyntaxWarning: invalid escape sequence '\V'
<>:6: SyntaxWarning: invalid escape sequence '\V'
C:\Users\stefa\AppData\Local\Temp\ipykernel_19844\1615316851.py:3: SyntaxWarning: invalid escape sequence '\V'
  classifier_3_path = "C:\VScode Git\DS2\Assignment II\stanford-ner-2020-11-17\classifiers\english.all.3class.distsim.crf.ser.gz"
C:\Users\stefa\AppData\Local\Temp\ipykernel_19844\1615316851.py:4: SyntaxWarning: invalid escape sequence '\V'
  classifier_4_path = "C:\VScode Git\DS2\Assignment II\stanford-ner-2020-11-17\classifiers\english.conll.4class.distsim.crf.ser.gz"
C:\Users\stefa\AppData\Local\Temp\ipykernel_19844\1615316851.py:5: SyntaxWarning: invalid escape sequence '\V'
  classifier_7_path = "C:\VScode Git\DS2\Assignment II\stanford-ner-2020-11-17\classifiers\english.muc.7class.distsim.crf.ser.gz"
C:\Users\stefa\AppData\Local\Temp\ipykernel_19844\1615316851.py:6: SyntaxWarning: invalid escape sequence '\V'
  jar_path = "C:\VScode Git\DS2\Assignment II\stanford-ner-2020-11-17\stanford-ner.jar"
```

3-class tagger:

[ 'William', 'Shakespeare', 'Barnardo', 'Horatio', 'Horatio', 'Barnardo', 'Holla', 'Barnardo', 'Horatio', 'Horatio', 'Barnardo', 'Marcellus', 'Heauen', 'Motiue', 'Cocke', 'Cocke', 'Obiect', 'Sauiours', 'Birch', 'Haile', 'Beene', 'Cocke', 'Soule', 'Growes', 'Morne', 'Doe', 'Primrose', 'Doe', 'Marke', 'Angell', 'Queene', 'Booke', 'Horatio', 'Hamlet', 'Guildensterne', 'Maiestie', 'Breuitie', 'Soule', 'Sonne', 'Sunne', 'Maiestie', 'Hamlet', 'Honestie', 'Madnesse', 'Sanitie', 'Rosincran', 'Begger', 'Lenton', 'Foyle', 'Doe', 'Boyes', 'Philosophie', 'Winde', 'Hawke', 'Hearke', 'Rossius', 'Iephta', 'Sable', 'Armes', 'Blacke', 'Pyrrhus', 'Stoopes', 'Pyrrhus', 'Milkie', 'Orbe', 'Pyrrhus', 'Pyrrhus', 'Naue', 'Bisson', 'Pyrrhus', 'Lye', 'Sonne', 'Diuell', 'Gertrude', 'Rich', 'Beautie', 'Beautie', 'Fellowes', 'Rose', 'Honie', 'Melancholly', 'Queene', 'Mother', 'Bodie', 'Soule', 'Mooones', 'Hymen', 'Feare', 'Willes', 'Rauen', 'Damon', 'Queene', 'Soule', 'Soule', 'Armour', 'Fetters', 'Raine', 'Ile', 'Sonne', 'Sonne', 'Begger', 'Mother', 'Maiesty', 'Cocke', 'Sonne', 'Burne', 'Steward', 'Fennell', 'Rew', 'Robin', 'Fellowes', 'Fellowes', 'Rosincrance', 'Wood', 'Norman', 'Laer.', 'Vpon', 'Hamlet', 'Laertes', 'Simples', 'Gallowes', 'Horatio', 'Cudgell', 'Asse', 'Scull', 'Sheepe', 'Calues', 'Hamlet', 'Tanner', 'Fellowes', 'Scull', 'Scull', 'Alexander', 'Alexander', 'died:', 'Alexander', 'Alexander', 'Akers', 'Dogge', 'Horatio', 'Gertrude', 'Yeomans', 'Assis', 'Modell', 'Bonet', 'Germaine', 'Canon', 'Complie', 'Hamlet', 'Hamlet', 'King.', 'Giue', 'Antike', 'Roman', 'Heauen', 'Ile', 'Guildensterne', 'Peale']

158

4-class tagger:

[ 'William', 'Shakespeare', 'Scoena', 'Barnardo', 'Haue', 'Horatio', 'Horatio', 'Dane', 'Fran.', 'Giue', 'Barnardo', 'Holla', 'Barnardo', 'Horatio', 'Marcellus', 'Horatio', 'Barnardo', 'Marcellus', 'Bell', 'Warlike', 'Armour', 'Norway', 'Pollax', 'Martiall', 'Taske', 'Motiue', 'Romage', 'Marcellus', 'Maiesticall', 'Mockery', 'Barn.', 'Cocke', 'Cocke', 'Throate', 'Awake', 'Obiect', 'Sauiours', 'Birch', 'Walkes', 'Claudius', 'King', 'Gertrude', 'Ioyntrousse', 'Dole', 'Taken', 'Colleagued', 'Voltemand', 'Vncle', 'Giuing', 'Voltemand', 'Head', 'Natie', 'Dread', 'Denmarke', 'Seeke', 'Inky', 'Cloake', 'Riuer', 'Minde', 'An', 'Vnderstanding', 'Sonne', 'Madam', 'Madam', 'Denmarke', 'Manet', 'Hiperion', 'Haile', 'friend', 'Ile', 'Marcellus', 'Fathers', 'Funerall', 'Heauens', 'Beene', 'Ielly', 'Cocke', 'His', 'Beward', 'Hell', 'twelue', 'Ile', 'Conuoy', 'Violet', 'Soule', 'Growes', 'Prodigall', 'Doe', 'Pastors', 'doe', 'Shew', 'Grapple', 'Tis', 'Tis', 'Vnsifted', 'Doe', 'Marry', 'Ile', 'Prodigall', 'Soule', 'Giues', 'Doe', 'Keepes', 'Royall', 'Hearsed', 'Sepulcher', 'Wherein', 'Marble', 'Reuisits', 'Doe', 'Vnhand', 'Heau'n', 'Ile', 'Ghost', 'Ile', 'Marke', 'Reuenge', 'Traitorous', 'Lewdnesse', 'Angell', 'Vncle', 'Gates', 'Aygre', 'Tetter', 'Queene', 'Soule', 'Matine', 'Booke', 'Horatio', 'Hamlet', 'Hor.', 'Heauen', 'Infaieth', 'Armes', 'God', 'Reynoldo', 'Danskens', 'God', 'Polon.', 'Obserue', 'Hamlet', 'Rosincrance', 'Hamlet', 'Heauens', 'Whereon', 'Giues', 'Maiestie', 'Dutie', 'Breuitie', 'Soule', 'Sonne', 'Dutie', 'Soules', 'Madam', 'Sunne', 'Maiestie', 'Queene', 'Deske', 'Hamlet', 'Ile', 'Magots', 'Honestie', 'Sanitie', 'Rosincran', 'God', 'Dungeons', 'Denmarke', 'Begger', 'Maiesticall', 'Roofe', 'Lenton', 'Knight', 'Boyes', 'Hercules', 'Vnckle', 'Ducates', 'Philosophie', 'Aunt', 'Mother', 'Winde', 'Hawke', 'Mark', 'Rossius', 'Plautus', 'Iephta', 'Iudge', 'Iephta', 'God', 'Pons', 'Chanson', 'Mistress', 'Byrlady', 'Ladiship', 'God', 'Cauiarie', 'Sallets', 'Pyrrhus', 'Pyrrhus', 'Sable', 'Armes', 'Blacke', 'Pyrrhus', 'Olde', 'Grandsire', 'Priam', 'Anon', 'Rebellious', 'Pyrrhus', 'Priam', 'Stoopes', 'Pyrrhus', 'Milkie', 'Tyrant', 'Pyrrhus', 'Newtrall', 'Orbe', 'Anon', 'Pyrrhus', 'Pyrrhus', 'Queene', 'Bisson', 'Pyrrhus', 'Gods', 'Dost', 'Manet', 'God', 'Hecuba', 'Tweakes', 'Lye', 'Gall', 'Slaues', 'Asse', 'Sonne', 'Deere', 'Vnkle.', 'Ile', 'Ile', 'Ile', 'Wherein', 'Ile', 'Lunacy', 'Rosin.', 'Madnesse', 'Gertrude', 'Vertues', 'Will', 'Deuotions', 'Harlots', 'Cheeke', 'Deuoutly', 'Scornes', 'Lawes', 'Honestie', 'Beautie', 'Fellowes', 'Heauen', 'Knaues', 'Farwell', 'God', 'Gods', 'Honie', 'Bels', 'Lord', 'Hamlet', 'Bodie', 'Hindges', 'Dost', 'Soule', 'Mistress', 'Acte', 'Ayre', 'Iulius', 'Brutus', 'Hoboyes', 'Layes', 'Anon', 'Queene', 'Poysoner', 'Wooes', 'Patientie', 'Poesie', 'Mooones', 'Hymen', 'Vnite', 'Sunne', 'Moone', 'Make', 'Feare', 'Wormwood', 'Fruite', 'Greefe', 'Greefe', 'Willes', 'Gonzago', 'Dukes', 'Lucianus', 'Rauen', 'Drugges', 'Murtherer', 'Deere', 'Hart', 'Turke', 'Prouinciall', 'Roses', 'Damon', 'deere', 'This', 'Realme', 'Paiocke', 'Rosincrance', 'Musicke', 'Guild.', 'Good', 'Dutie', 'Ventiges', 'God', 'Verie', 'Hell', 'Soule', 'Soule', 'Maiestie', 'Armour', 'Somnet', 'Fetters', 'Brothers', 'Raine', 'Whereeto', 'Forgiue', 'Buyes', 'Soule', 'Physicke', 'him', 'Ile', 'Ile', 'Aye', 'Hyperions', 'God', 'Ardure', 'Tinct', 'Ham.', 'Stye', 'Sonne', 'Mother', 'Soldiours', 'Forgiue', 'Scourge', 'Pinch', 'Wanton', 'Maiesty', 'Hamlet', 'Mother', 'Clossets', 'Hamlet', 'Kinne', 'Chappell', 'Sonne', 'King', 'Deliberate', 'Begger', 'Progresse', 'K.', 'Hamlet', 'Barke', 'Mother', 'Mother', 'Fortinbras', 'Claimes', 'Queene', 'Ophelia', 'Maiesty', 'God', 'God', 'Ophelia', 'Cocke', 'Goodnight', 'Ladies', 'Goodnight', 'Sonne', 'Ophelia', 'Diuided', 'Diuinity', 'dead?', 'Ile', 'Ile', 'Looser', 'Laer.', 'None', 'Burne', 'Steward', 'Fennell', 'Rew', 'Rew', 'Robin', 'Ally', 'Flaxen', 'Gramercy', 'God', 'Hatchment', 'Saylor', 'God', 'Fellowes', 'Warlicke', 'Theeues', 'Fellowes', 'Rosincrance', 'Norman', 'Norman', 'Laer.', 'Vpon', 'Lamound', 'Hamlet', 'Hamlet', 'Hamlet', 'Ile', 'drinke', 'Ile', 'Clambring', 'Aduer', 'Graue', 'Goodman', 'Deuer', 'Ditchers', 'Armes', 'Adam', 'dig'd;', 'Hamlet', 'Horatio', 'Cudgell', 'Asse', 'Clowne', 'Scull', 'Caines', 'Asse', 'Good', 'Morrow', 'Mazard', 'Clowne', 'Pickhaxe', 'Scull', 'Hamlet', 'Tanner', 'Decayer', 'Fellowes', 'Scull', 'Scull', 'Yoricks', 'Alexander', 'Alexander', 'died:', 'Alexander', 'Alexander', 'Cerimony', 'Buriall', 'Laer.', 'Must', 'Soules', 'Angell', 'Graue', 'Dane', 'Laer.', 'Vntill', 'God', 'Make', 'Ossa', 'Hercules', 'Dogge', 'Horatio', 'Gertrude', 'Graue', 'Horatio', 'Ham.', 'Denmarks', 'Goblins', 'Yeomans', 'Coniuration', 'Garland', 'Assis', 'Heauen', 'Modell', 'Popt', 'Denmarke', 'Bonet', 'Maiesty', 'Germaine', 'Lapwing', 'Complie', 'Hamlet', 'Hamlet', 'Hamlet', 'Hamlet', 'Hamlet', 'Arrow', 'Giue', 'Cousen', 'Foyles', 'Hamlet', 'Richer', 'Ham.', 'Ile', 'Sonne', 'Queene', 'Carowes', 'Madam', 'Queene', 'Woodcocke', 'Vnbated', 'Queene', 'Goodnight', 'Rosincrance', 'Guildensterne', 'Inuite']

512

7-class tagger:

[ 'Horatio', 'Dane', 'Fran.', 'Giue', 'Barnardo', 'Horatio', 'Barnardo', 'Marcellus', 'Horatio', 'Barn.', 'Lookes', 'King?', 'Marke', 'Armour', 'Martiall', 'Cocke', 'Cocke', 'Sauiours', 'Birch', 'Hiperion', 'King', 'Heauens', 'Beene', 'Cocke', 'Doe', 'Soule', 'Growes', 'Doe', 'Shew', 'Primrose', 'Doe', 'Lord', 'Hamlet', 'Beleeue', 'Doe', 'Royall', 'Doe', 'Marke', 'Reuenge', 'Angell', 'Soule', 'Booke', 'Neuer', 'Giue', 'Doe', 'Marke', 'Thanks', 'Guildensterne', 'Heauens', 'Breuitie', 'Soule', 'Sunne', 'Lord', 'Hamlet', 'Marke', 'Madnesse', 'Lenton', 'Doe', 'Boyes', 'Vnckle', 'Winde', 'Hawke', 'Sable', 'Armes', 'Blacke', 'Pyrrhus', 'Milkie', 'Tyrant', 'Pyrrhus', 'Pyrrhus', 'Bisson', 'Sonne', 'Diuell', 'Gertrude', 'leane', 'Naturall', 'Lawes', 'Fellowes', 'Bodie', 'deere', 'Soule', 'Belike', 'Hymen', 'Sunne', 'Moone', 'Feare', 'Greefe', 'Willes', 'Sleepe', 'rocke', 'Rauen', 'Giue', 'Deere', 'Hart', 'Damon', 'deere', 'This', 'Realme', 'King', 'Verie', 'Soule', 'Soule', 'Armour', 'Fetters', 'Raine', 'Soule', 'Aye', 'Sonne', 'Chappell', 'Sonne', 'King', 'leane', 'Begger', 'Cocke', 'Sonne', 'Iudgement', 'pierce', 'Burne', 'Steward', 'Fennell', 'Rew', 'Rew', 'Robin', 'Colaterall', 'Fellowes', 'Fellowes', 'Breake', 'Norman', 'was't?', 'Kin.', 'A', 'Norman', 'Laer.', 'Vpon', 'Laertes', 'Graue', 'Adam', 'dig'd;', 'Gallowes', 'Cudgell', 'Custom', 'Scull', 'Morrow', 'Scull', 'Vpon', 'Tanner', 'Scull', 'Scull', 'Alexander', 'died:', 'Alexander', 'Alexander', 'Soules', 'Ministring', 'Angell', 'Graue', 'Dane', 'Laer.', 'Heare', 'Dogge', 'Horatio', 'Gertrude', 'Graue', 'Yeomans', 'Bonet', 'Germaine', 'King', 'Giue', 'Giue', 'Verie', 'Giue', 'Sonne', 'Heere', 'Rosincrance', 'Guildensterne', 'Captaines', 'Beare', 'Hamlet']

178