```
{
 "cells": [
  {
   "cell_type": "code",
   "execution_count": 4,
   "id": "0d159180",
   "metadata": {},
   "outputs": [],
   "source": [
    "import nltk\n",
    "import os"
   ]
  },
  {
   "cell_type": "code",
   "execution_count": 13,
   "id": "f427fb44",
   "metadata": {},
   "outputs": [
    {
     "name": "stdout",
     "output_type": "stream",
     "text": [
      "Downloading 'averaged_perceptron_tagger'...\n"
     ]
    },
    {
     "name": "stderr",
     "output_type": "stream",
     "text": [
      "[nltk_data] Downloading package averaged_perceptron_tagger to\n",
      "[nltk_data]
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n"
     ]
    },
    {
     "name": "stdout",
     "output_type": "stream",
     "text": [
      "Download complete.\n"
     ]
    },
    {
     "name": "stderr",
     "output_type": "stream",
     "text": [
      "[nltk_data]   Unzipping
taggers\\averaged_perceptron_tagger.zip.\n"
     ]
    }
   ],
   "source": [
    "try:\n",
    "    nltk.data.find('taggers/averaged_perceptron_tagger')\n",
    "except LookupError:\n",
    "    print(\"Downloading 'averaged_perceptron_tagger'...\")\n",
    "    nltk.download('averaged_perceptron_tagger')\n",
    "    print(\"Download complete.\")"
```

```
 ]
},
{
 "cell_type": "code",
 "execution_count": 5,
 "id": "5d5cf4c1",
 "metadata": {},
 "outputs": [],
 "source": [
  "import nltk.corpus"
 ]
},
{
 "cell_type": "code",
 "execution_count": 6,
 "id": "0d6b4706",
 "metadata": {},
 "outputs": [],
 "source": [
  "from nltk.tokenize import word_tokenize\n",
  "from nltk.tokenize import RegexpTokenizer"
 ]
},
{
 "cell_type": "code",
 "execution_count": 7,
 "id": "2d50c89d",
 "metadata": {},
 "outputs": [],
 "source": [
  "from nltk.data import load"
 ]
},
{
 "cell_type": "code",
 "execution_count": 8,
 "id": "5e28dded",
 "metadata": {},
 "outputs": [],
 "source": [
  "sent = \"Mary is driving a big car.\""
 ]
},
{
 "cell_type": "code",
 "execution_count": 10,
 "id": "5341e9e4",
 "metadata": {},
 "outputs": [],
 "source": [
  "sent_tokens = word_tokenize(sent)"
 ]
},
{
 "cell_type": "code",
 "execution_count": 14,
 "id": "841d76e1",
 "metadata": {},
```

```
 "outputs": [
  {
   "name": "stdout",
   "output_type": "stream",
   "text": [
    "[('Mary', 'NNP')]\n",
    "[('is', 'VBZ')]\n",
    "[('driving', 'VBG')]\n",
    "[('a', 'DT')]\n",
    "[('big', 'JJ')]\n",
    "[('car', 'NN')]\n",
    "[('.', '.')]\n"
   ]
  }
 ],
 "source": [
  "for token in sent_tokens:\n",
  "    print(nltk.pos_tag([token]))"
 ]
},
{
 "cell_type": "code",
 "execution_count": null,
 "id": "59595709",
 "metadata": {},
 "outputs": [],
 "source": [
  "import nltk\n",
  "from nltk.tokenize import word_tokenize\n"
 ]
},
{
 "cell_type": "code",
 "execution_count": 15,
 "id": "8aae81ef",
 "metadata": {},
 "outputs": [],
 "source": [
  "sent2 = \"John is eating a delicious cake\""
 ]
},
{
 "cell_type": "code",
 "execution_count": 17,
 "id": "d6f9dfab",
 "metadata": {},
 "outputs": [
  {
   "name": "stdout",
   "output_type": "stream",
   "text": [
    "[('John', 'NNP')]\n",
    "[('is', 'VBZ')]\n",
    "[('eating', 'VBG')]\n",
    "[('a', 'DT')]\n",
    "[('delicious', 'JJ')]\n",
    "[('cake', 'NN')]\n"
   ]
```

```
  }
 ],
 "source": [
  "sent2_tokens = word_tokenize(sent2)\n",
  "\n",
  "for token in sent2_tokens:\n",
  "    print(nltk.pos_tag([token]))"
 ]
},
{
 "cell_type": "code",
 "execution_count": 18,
 "id": "866ccfd3",
 "metadata": {},
 "outputs": [],
 "source": [
  "sent3 = \"Jim eats a banana\""
 ]
},
{
 "cell_type": "code",
 "execution_count": 19,
 "id": "10d1699f",
 "metadata": {},
 "outputs": [
  {
   "name": "stdout",
   "output_type": "stream",
   "text": [
    "[('Jim', 'NNP')]\n",
    "[('eats', 'NNS')]\n",
    "[('a', 'DT')]\n",
    "[('banana', 'NN')]\n"
   ]
  }
 ],
 "source": [
  "sent3_tokens = word_tokenize(sent3)\n",
  "\n",
  "for tokens in sent3_tokens:\n",
  "    print(nltk.pos_tag([tokens]))"
 ]
},
{
 "cell_type": "code",
 "execution_count": 20,
 "id": "3cf0604e",
 "metadata": {},
 "outputs": [],
 "source": [
  "reg_tokenizer = RegexpTokenizer('(?u)\\w+|\\$[\\d\\.]+|\\S+')"
 ]
},
{
 "cell_type": "code",
 "execution_count": 21,
 "id": "de8e0aa0",
 "metadata": {},
```

```
 "outputs": [],
 "source": [
  "regex_tokenize = reg_tokenizer.tokenize(sent3)"
 ]
},
{
 "cell_type": "code",
 "execution_count": 25,
 "id": "6ef7a908",
 "metadata": {},
 "outputs": [
  {
   "data": {
    "text/plain": [
     "['Jim', 'eats', 'a', 'banana']"
    ]
   },
   "execution_count": 25,
   "metadata": {},
   "output_type": "execute_result"
  }
 ],
 "source": [
  "regex_tokenize"
 ]
},
{
 "cell_type": "code",
 "execution_count": 26,
 "id": "597005b0",
 "metadata": {},
 "outputs": [],
 "source": [
  "regex_tag = nltk.pos_tag(regex_tokenize)"
 ]
},
{
 "cell_type": "code",
 "execution_count": 27,
 "id": "ff19199d",
 "metadata": {},
 "outputs": [
  {
   "data": {
    "text/plain": [
     "[('Jim', 'NNP'), ('eats', 'VBZ'), ('a', 'DT'), ('banana', 'NN')]"
    ]
   },
   "execution_count": 27,
   "metadata": {},
   "output_type": "execute_result"
  }
 ],
 "source": [
  "regex_tag"
 ]
},
{
```

```
   "cell_type": "code",
   "execution_count": 29,
   "id": "ae1cfad3",
   "metadata": {},
   "outputs": [
    {
     "name": "stdout",
     "output_type": "stream",
     "text": [
      "NLTK data (including corpora) not found. Downloading 'popular'
bundle...\n"
     ]
    },
    {
     "name": "stderr",
     "output_type": "stream",
     "text": [
      "[nltk_data] Downloading collection 'popular'\n",
      "[nltk_data]    | \n",
      "[nltk_data]    | Downloading package cmudict to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\cmudict.zip.\n",
      "[nltk_data]    | Downloading package gazetteers to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\gazetteers.zip.\n",
      "[nltk_data]    | Downloading package genesis to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\genesis.zip.\n",
      "[nltk_data]    | Downloading package gutenberg to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\gutenberg.zip.\n",
      "[nltk_data]    | Downloading package inaugural to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\inaugural.zip.\n",
      "[nltk_data]    | Downloading package movie_reviews to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\movie_reviews.zip.\n",
      "[nltk_data]    | Downloading package names to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\names.zip.\n",
      "[nltk_data]    | Downloading package shakespeare to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\shakespeare.zip.\n",
      "[nltk_data]    | Downloading package stopwords to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\stopwords.zip.\n",
      "[nltk_data]    | Downloading package treebank to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
```

```
      "[nltk_data]    |   Unzipping corpora\\treebank.zip.\n",
      "[nltk_data]    | Downloading package twitter_samples to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\twitter_samples.zip.\n",
      "[nltk_data]    | Downloading package omw to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    | Downloading package omw-1.4 to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    | Downloading package wordnet to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    | Downloading package wordnet2021 to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    | Downloading package wordnet31 to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    | Downloading package wordnet_ic to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\wordnet_ic.zip.\n",
      "[nltk_data]    | Downloading package words to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping corpora\\words.zip.\n",
      "[nltk_data]    | Downloading package maxent_ne_chunker to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Unzipping chunkers\\maxent_ne_chunker.zip.\n",
      "[nltk_data]    | Downloading package punkt to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Package punkt is already up-to-date!\n",
      "[nltk_data]    | Downloading package snowball_data to\n",
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n"
     ]
    },
    {
     "name": "stdout",
     "output_type": "stream",
     "text": [
      "Download complete.\n",
      "The 'corpora' directory was not found at
C:\\Users\\lavan/nltk_data\\corpora. Ensure NLTK data is correctly
downloaded and installed.\n"
     ]
    },
    {
     "name": "stderr",
     "output_type": "stream",
     "text": [
      "[nltk_data]    | Downloading package averaged_perceptron_tagger
to\n",
```

```
      "[nltk_data]    |
C:\\Users\\lavan\\AppData\\Roaming\\nltk_data...\n",
      "[nltk_data]    |   Package averaged_perceptron_tagger is already
up-\n",
      "[nltk_data]    |       to-date!\n",
      "[nltk_data]    | \n",
      "[nltk_data]  Done downloading collection popular\n"
     ]
    }
   ],
   "source": [
    "import nltk\n",
    "import os\n",
    "\n",
    "try:\n",
    "    nltk.data.find('corpora/wordnet.zip') \n",
    "except LookupError:\n",
    "    print(\"NLTK data (including corpora) not found. Downloading
'popular' bundle...\")\n",
    "    nltk.download('popular')\n",
    "    print(\"Download complete.\")\n",
    "\n",
    "nltk_data_path = nltk.data.path[0]\n",
    "\n",
    "corpora_dir = os.path.join(nltk_data_path, \"corpora\")\n",
    "\n",
    "if os.path.isdir(corpora_dir):\n",
    "    print(f\"Contents of '{corpora_dir}':\")\n",
    "    print(os.listdir(corpora_dir))\n",
    "else:\n",
    "    print(f\"The 'corpora' directory was not found at {corpora_dir}.
\"\n",
    "          \"Ensure NLTK data is correctly downloaded and
installed.\")"
   ]
  },
  {
   "cell_type": "code",
   "execution_count": 31,
   "id": "66e223a0",
   "metadata": {},
   "outputs": [
    {
     "name": "stdout",
     "output_type": "stream",
     "text": [
      "['cmudict', 'cmudict.zip', 'gazetteers', 'gazetteers.zip',
'genesis', 'genesis.zip', 'gutenberg', 'gutenberg.zip', 'inaugural',
'inaugural.zip', 'movie_reviews', 'movie_reviews.zip', 'names',
'names.zip', 'omw-1.4.zip', 'omw.zip', 'shakespeare', 'shakespeare.zip',
'stopwords', 'stopwords.zip', 'treebank', 'treebank.zip',
'twitter_samples', 'twitter_samples.zip', 'wordnet.zip',
'wordnet2021.zip', 'wordnet31.zip', 'wordnet_ic', 'wordnet_ic.zip',
'words', 'words.zip']\n"
     ]
    }
   ],
   "source": [
```

```
  "print(os.listdir(nltk.data.find(\"corpora\")))"
 ]
},
{
 "cell_type": "code",
 "execution_count": 32,
 "id": "dac3ffdf",
 "metadata": {},
 "outputs": [
  {
   "data": {
    "text/plain": [
     "['austen-emma.txt',\n",
     " 'austen-persuasion.txt',\n",
     " 'austen-sense.txt',\n",
     " 'bible-kjv.txt',\n",
     " 'blake-poems.txt',\n",
     " 'bryant-stories.txt',\n",
     " 'burgess-busterbrown.txt',\n",
     " 'carroll-alice.txt',\n",
     " 'chesterton-ball.txt',\n",
     " 'chesterton-brown.txt',\n",
     " 'chesterton-thursday.txt',\n",
     " 'edgeworth-parents.txt',\n",
     " 'melville-moby_dick.txt',\n",
     " 'milton-paradise.txt',\n",
     " 'shakespeare-caesar.txt',\n",
     " 'shakespeare-hamlet.txt',\n",
     " 'shakespeare-macbeth.txt',\n",
     " 'whitman-leaves.txt']"
    ]
   },
   "execution_count": 32,
   "metadata": {},
   "output_type": "execute_result"
  }
 ],
 "source": [
  "nltk.corpus.gutenberg.fileids()"
 ]
},
{
 "cell_type": "code",
 "execution_count": 33,
 "id": "71be5db6",
 "metadata": {},
 "outputs": [],
 "source": [
  "hamlet=nltk.corpus.gutenberg.words('shakespeare-hamlet.txt')"
 ]
},
{
 "cell_type": "code",
 "execution_count": 34,
 "id": "c4d6f635",
 "metadata": {},
 "outputs": [
  {
```

```
   "data": {
    "text/plain": [
     "37360"
    ]
   },
   "execution_count": 34,
   "metadata": {},
   "output_type": "execute_result"
  }
 ],
 "source": [
  "len(hamlet)"
 ]
},
{
 "cell_type": "code",
 "execution_count": 35,
 "id": "b24836d8",
 "metadata": {},
 "outputs": [],
 "source": [
  "hamlet_pos = []"
 ]
},
{
 "cell_type": "code",
 "execution_count": 36,
 "id": "f27602aa",
 "metadata": {},
 "outputs": [],
 "source": [
  "for word in hamlet[:2000]:\n",
  "    word_pos=nltk.pos_tag([word])\n",
  "    hamlet_pos.append(word_pos)"
 ]
},
{
 "cell_type": "code",
 "execution_count": 37,
 "id": "5d24f5c2",
 "metadata": {},
 "outputs": [
  {
   "data": {
    "text/plain": [
     "[[('[', 'NN')],\n",
     " [('The', 'DT')],\n",
     " [('Tragedie', 'NN')],\n",
     " [('of', 'IN')],\n",
     " [('Hamlet', 'NN')],\n",
     " [('by', 'IN')],\n",
     " [('William', 'NNP')],\n",
     " [('Shakespeare', 'NN')],\n",
     " [('1599', 'CD')],\n",
     " [(']', 'NN')],\n",
     " [('Actus', 'NN')],\n",
     " [('Primus', 'NN')],\n",
     " [('.', '.')],\n",
```

```
" [('Scoena', 'NN')],\n",
" [('Prima', 'NN')],\n",
" [('.', '.')],\n",
" [('Enter', 'NN')],\n",
" [('Barnardo', 'NN')],\n",
" [('and', 'CC')],\n",
" [('Francisco', 'NNP')],\n",
" [('two', 'CD')],\n",
" [('Centinels', 'NNS')],\n",
" [('.', '.')],\n",
" [('Barnardo', 'NN')],\n",
" [('.', '.')],\n",
" [('Who', 'WP')],\n",
" [(\"'\", \"''\")],\n",
" [('s', 'NN')],\n",
" [('there', 'RB')],\n",
" [('?', '.')],\n",
" [('Fran', 'NN')],\n",
" [('.', '.')],\n",
" [('Nay', 'NN')],\n",
" [('answer', 'NN')],\n",
" [('me', 'PRP')],\n",
" [(':', ':')],\n",
" [('Stand', 'NN')],\n",
" [('&', 'CC')],\n",
" [('vnfold', 'NN')],\n",
" [('your', 'PRP$')],\n",
" [('selfe', 'NN')],\n",
" [('Bar', 'NN')],\n",
" [('.', '.')],\n",
" [('Long', 'RB')],\n",
" [('liue', 'NN')],\n",
" [('the', 'DT')],\n",
" [('King', 'VBG')],\n",
" [('Fran', 'NN')],\n",
" [('.', '.')],\n",
" [('Barnardo', 'NN')],\n",
" [('?', '.')],\n",
" [('Bar', 'NN')],\n",
" [('.', '.')],\n",
" [('He', 'PRP')],\n",
" [('Fran', 'NN')],\n",
" [('.', '.')],\n",
" [('You', 'PRP')],\n",
" [('come', 'VB')],\n",
" [('most', 'JJS')],\n",
" [('carefully', 'RB')],\n",
" [('vpon', 'NN')],\n",
" [('your', 'PRP$')],\n",
" [('houre', 'NN')],\n",
" [('Bar', 'NN')],\n",
" [('.', '.')],\n",
" [(\"'\", \"''\")],\n",
" [('Tis', 'NN')],\n",
" [('now', 'RB')],\n",
" [('strook', 'NN')],\n",
" [('twelue', 'NN')],\n",
" [(',', ',')],\n",
```

```
" [('get', 'VB')],\n",
" [('thee', 'NN')],\n",
" [('to', 'TO')],\n",
" [('bed', 'NN')],\n",
" [('Francisco', 'NNP')],\n",
" [('Fran', 'NN')],\n",
" [('.', '.')],\n",
" [('For', 'IN')],\n",
" [('this', 'DT')],\n",
" [('releefe', 'NN')],\n",
" [('much', 'JJ')],\n",
" [('thankes', 'NNS')],\n",
" [(':', ':')],\n",
" [(\"'\", \"''\")],\n",
" [('Tis', 'NN')],\n",
" [('bitter', 'NN')],\n",
" [('cold', 'NN')],\n",
" [(',', ',')],\n",
" [('And', 'CC')],\n",
" [('I', 'PRP')],\n",
" [('am', 'VBP')],\n",
" [('sicke', 'NN')],\n",
" [('at', 'IN')],\n",
" [('heart', 'NN')],\n",
" [('Barn', 'NN')],\n",
" [('.', '.')],\n",
" [('Haue', 'NN')],\n",
" [('you', 'PRP')],\n",
" [('had', 'VBD')],\n",
" [('quiet', 'JJ')],\n",
" [('Guard', 'NN')],\n",
" [('?', '.')],\n",
" [('Fran', 'NN')],\n",
" [('.', '.')],\n",
" [('Not', 'RB')],\n",
" [('a', 'DT')],\n",
" [('Mouse', 'NN')],\n",
" [('stirring', 'VBG')],\n",
" [('Barn', 'NN')],\n",
" [('.', '.')],\n",
" [('Well', 'RB')],\n",
" [(',', ',')],\n",
" [('goodnight', 'NN')],\n",
" [('.', '.')],\n",
" [('If', 'IN')],\n",
" [('you', 'PRP')],\n",
" [('do', 'VB')],\n",
" [('meet', 'NN')],\n",
" [('Horatio', 'NN')],\n",
" [('and', 'CC')],\n",
" [('Marcellus', 'NN')],\n",
" [(',', ',')],\n",
" [('the', 'DT')],\n",
" [('Riuals', 'NNS')],\n",
" [('of', 'IN')],\n",
" [('my', 'PRP$')],\n",
" [('Watch', 'NN')],\n",
" [(',', ',')],\n",
```

```
" [('bid', 'NN')],\n",
" [('them', 'PRP')],\n",
" [('make', 'VB')],\n",
" [('hast', 'NN')],\n",
" [('.', '.')],\n",
" [('Enter', 'NN')],\n",
" [('Horatio', 'NN')],\n",
" [('and', 'CC')],\n",
" [('Marcellus', 'NN')],\n",
" [('.', '.')],\n",
" [('Fran', 'NN')],\n",
" [('.', '.')],\n",
" [('I', 'PRP')],\n",
" [('thinke', 'NN')],\n",
" [('I', 'PRP')],\n",
" [('heare', 'NN')],\n",
" [('them', 'PRP')],\n",
" [('.', '.')],\n",
" [('Stand', 'NN')],\n",
" [(':', ':')],\n",
" [('who', 'WP')],\n",
" [(\"'\", \"''\")],\n",
" [('s', 'NN')],\n",
" [('there', 'RB')],\n",
" [('?', '.')],\n",
" [('Hor', 'NN')],\n",
" [('.', '.')],\n",
" [('Friends', 'NNS')],\n",
" [('to', 'TO')],\n",
" [('this', 'DT')],\n",
" [('ground', 'NN')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [('And', 'CC')],\n",
" [('Leige', 'NN')],\n",
" [('-', ':')],\n",
" [('men', 'NNS')],\n",
" [('to', 'TO')],\n",
" [('the', 'DT')],\n",
" [('Dane', 'NN')],\n",
" [('Fran', 'NN')],\n",
" [('.', '.')],\n",
" [('Giue', 'NN')],\n",
" [('you', 'PRP')],\n",
" [('good', 'JJ')],\n",
" [('night', 'NN')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [('O', 'NN')],\n",
" [('farwel', 'NN')],\n",
" [('honest', 'NN')],\n",
" [('Soldier', 'NN')],\n",
" [(',', ',')],\n",
" [('who', 'WP')],\n",
" [('hath', 'NN')],\n",
" [('relieu', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('d', 'NN')],\n",
```

```
" [('you', 'PRP')],\n",
" [('?', '.')],\n",
" [('Fra', 'NN')],\n",
" [('.', '.')],\n",
" [('Barnardo', 'NN')],\n",
" [('ha', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('s', 'NN')],\n",
" [('my', 'PRP$')],\n",
" [('place', 'NN')],\n",
" [(':', ':')],\n",
" [('giue', 'NN')],\n",
" [('you', 'PRP')],\n",
" [('goodnight', 'NN')],\n",
" [('.', '.')],\n",
" [('Exit', 'NN')],\n",
" [('Fran', 'NN')],\n",
" [('.', '.')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [('Holla', 'NN')],\n",
" [('Barnardo', 'NN')],\n",
" [('Bar', 'NN')],\n",
" [('.', '.')],\n",
" [('Say', 'NN')],\n",
" [(',', ',')],\n",
" [('what', 'WP')],\n",
" [('is', 'VBZ')],\n",
" [('Horatio', 'NN')],\n",
" [('there', 'RB')],\n",
" [('?', '.')],\n",
" [('Hor', 'NN')],\n",
" [('.', '.')],\n",
" [('A', 'DT')],\n",
" [('peece', 'NN')],\n",
" [('of', 'IN')],\n",
" [('him', 'PRP')],\n",
" [('Bar', 'NN')],\n",
" [('.', '.')],\n",
" [('Welcome', 'VB')],\n",
" [('Horatio', 'NN')],\n",
" [(',', ',')],\n",
" [('welcome', 'NN')],\n",
" [('good', 'JJ')],\n",
" [('Marcellus', 'NN')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [('What', 'WP')],\n",
" [(',', ',')],\n",
" [('ha', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('s', 'NN')],\n",
" [('this', 'DT')],\n",
" [('thing', 'NN')],\n",
" [('appear', 'VB')],\n",
" [(\"'\", \"''\")],\n",
" [('d', 'NN')],\n",
" [('againe', 'NN')],\n",
```

```
" [('to', 'TO')],\n",
" [('night', 'NN')],\n",
" [('Bar', 'NN')],\n",
" [('.', '.')],\n",
" [('I', 'PRP')],\n",
" [('haue', 'NN')],\n",
" [('seene', 'NN')],\n",
" [('nothing', 'NN')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [('Horatio', 'NN')],\n",
" [('saies', 'NNS')],\n",
" [(',', ',')],\n",
" [(\"'\", \"''\")],\n",
" [('tis', 'NN')],\n",
" [('but', 'CC')],\n",
" [('our', 'PRP$')],\n",
" [('Fantasie', 'NN')],\n",
" [(',', ',')],\n",
" [('And', 'CC')],\n",
" [('will', 'MD')],\n",
" [('not', 'RB')],\n",
" [('let', 'VB')],\n",
" [('beleefe', 'NN')],\n",
" [('take', 'VB')],\n",
" [('hold', 'NN')],\n",
" [('of', 'IN')],\n",
" [('him', 'PRP')],\n",
" [('Touching', 'VBG')],\n",
" [('this', 'DT')],\n",
" [('dreaded', 'VBN')],\n",
" [('sight', 'NN')],\n",
" [(',', ',')],\n",
" [('twice', 'RB')],\n",
" [('seene', 'NN')],\n",
" [('of', 'IN')],\n",
" [('vs', 'NN')],\n",
" [(',', ',')],\n",
" [('Therefore', 'RB')],\n",
" [('I', 'PRP')],\n",
" [('haue', 'NN')],\n",
" [('intreated', 'VBN')],\n",
" [('him', 'PRP')],\n",
" [('along', 'IN')],\n",
" [('With', 'IN')],\n",
" [('vs', 'NN')],\n",
" [(',', ',')],\n",
" [('to', 'TO')],\n",
" [('watch', 'NN')],\n",
" [('the', 'DT')],\n",
" [('minutes', 'NNS')],\n",
" [('of', 'IN')],\n",
" [('this', 'DT')],\n",
" [('Night', 'NN')],\n",
" [(',', ',')],\n",
" [('That', 'DT')],\n",
" [('if', 'IN')],\n",
" [('againe', 'NN')],\n",
```

```
" [('this', 'DT')],\n",
" [('Apparition', 'NN')],\n",
" [('come', 'VB')],\n",
" [(',', ',')],\n",
" [('He', 'PRP')],\n",
" [('may', 'MD')],\n",
" [('approue', 'NN')],\n",
" [('our', 'PRP$')],\n",
" [('eyes', 'NNS')],\n",
" [(',', ',')],\n",
" [('and', 'CC')],\n",
" [('speake', 'NN')],\n",
" [('to', 'TO')],\n",
" [('it', 'PRP')],\n",
" [('Hor', 'NN')],\n",
" [('.', '.')],\n",
" [('Tush', 'NN')],\n",
" [(',', ',')],\n",
" [('tush', 'NN')],\n",
" [(',', ',')],\n",
" [(\"'\", \"''\")],\n",
" [('twill', 'NN')],\n",
" [('not', 'RB')],\n",
" [('appeare', 'NN')],\n",
" [('Bar', 'NN')],\n",
" [('.', '.')],\n",
" [('Sit', 'NN')],\n",
" [('downe', 'NN')],\n",
" [('a', 'DT')],\n",
" [('-', ':')],\n",
" [('while', 'IN')],\n",
" [(',', ',')],\n",
" [('And', 'CC')],\n",
" [('let', 'VB')],\n",
" [('vs', 'NN')],\n",
" [('once', 'RB')],\n",
" [('againe', 'NN')],\n",
" [('assaile', 'NN')],\n",
" [('your', 'PRP$')],\n",
" [('eares', 'NNS')],\n",
" [(',', ',')],\n",
" [('That', 'DT')],\n",
" [('are', 'VBP')],\n",
" [('so', 'RB')],\n",
" [('fortified', 'VBN')],\n",
" [('against', 'IN')],\n",
" [('our', 'PRP$')],\n",
" [('Story', 'NN')],\n",
" [(',', ',')],\n",
" [('What', 'WP')],\n",
" [('we', 'PRP')],\n",
" [('two', 'CD')],\n",
" [('Nights', 'NNS')],\n",
" [('haue', 'NN')],\n",
" [('seene', 'NN')],\n",
" [('Hor', 'NN')],\n",
" [('.', '.')],\n",
" [('Well', 'RB')],\n",
```

```
" [(',', ',')],\n",
" [('sit', 'NN')],\n",
" [('we', 'PRP')],\n",
" [('downe', 'NN')],\n",
" [(',', ',')],\n",
" [('And', 'CC')],\n",
" [('let', 'VB')],\n",
" [('vs', 'NN')],\n",
" [('heare', 'NN')],\n",
" [('Barnardo', 'NN')],\n",
" [('speake', 'NN')],\n",
" [('of', 'IN')],\n",
" [('this', 'DT')],\n",
" [('Barn', 'NN')],\n",
" [('.', '.')],\n",
" [('Last', 'JJ')],\n",
" [('night', 'NN')],\n",
" [('of', 'IN')],\n",
" [('all', 'DT')],\n",
" [(',', ',')],\n",
" [('When', 'WRB')],\n",
" [('yond', 'NN')],\n",
" [('same', 'JJ')],\n",
" [('Starre', 'NN')],\n",
" [('that', 'IN')],\n",
" [(\"'\", \"''\")],\n",
" [('s', 'NN')],\n",
" [('Westward', 'NNP')],\n",
" [('from', 'IN')],\n",
" [('the', 'DT')],\n",
" [('Pole', 'NN')],\n",
" [('Had', 'VBD')],\n",
" [('made', 'VBN')],\n",
" [('his', 'PRP$')],\n",
" [('course', 'NN')],\n",
" [('t', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('illume', 'NN')],\n",
" [('that', 'IN')],\n",
" [('part', 'NN')],\n",
" [('of', 'IN')],\n",
" [('Heauen', 'NN')],\n",
" [('Where', 'WRB')],\n",
" [('now', 'RB')],\n",
" [('it', 'PRP')],\n",
" [('burnes', 'NNS')],\n",
" [(',', ',')],\n",
" [('Marcellus', 'NN')],\n",
" [('and', 'CC')],\n",
" [('my', 'PRP$')],\n",
" [('selfe', 'NN')],\n",
" [(',', ',')],\n",
" [('The', 'DT')],\n",
" [('Bell', 'NNP')],\n",
" [('then', 'RB')],\n",
" [('beating', 'NN')],\n",
" [('one', 'CD')],\n",
" [('Mar', 'NN')],\n",
```

```
" [('.', '.')],\n",
" [('Peace', 'NN')],\n",
" [(',', ',')],\n",
" [('breake', 'NN')],\n",
" [('thee', 'NN')],\n",
" [('of', 'IN')],\n",
" [(':', ':')],\n",
" [('Enter', 'NN')],\n",
" [('the', 'DT')],\n",
" [('Ghost', 'NN')],\n",
" [('.', '.')],\n",
" [('Looke', 'NN')],\n",
" [('where', 'WRB')],\n",
" [('it', 'PRP')],\n",
" [('comes', 'VBZ')],\n",
" [('againe', 'NN')],\n",
" [('Barn', 'NN')],\n",
" [('.', '.')],\n",
" [('In', 'IN')],\n",
" [('the', 'DT')],\n",
" [('same', 'JJ')],\n",
" [('figure', 'NN')],\n",
" [(',', ',')],\n",
" [('like', 'IN')],\n",
" [('the', 'DT')],\n",
" [('King', 'VBG')],\n",
" [('that', 'IN')],\n",
" [(\"'\", \"''\")],\n",
" [('s', 'NN')],\n",
" [('dead', 'JJ')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [('Thou', 'NN')],\n",
" [('art', 'NN')],\n",
" [('a', 'DT')],\n",
" [('Scholler', 'NN')],\n",
" [(';', ':')],\n",
" [('speake', 'NN')],\n",
" [('to', 'TO')],\n",
" [('it', 'PRP')],\n",
" [('Horatio', 'NN')],\n",
" [('Barn', 'NN')],\n",
" [('.', '.')],\n",
" [('Lookes', 'NNS')],\n",
" [('it', 'PRP')],\n",
" [('not', 'RB')],\n",
" [('like', 'IN')],\n",
" [('the', 'DT')],\n",
" [('King', 'VBG')],\n",
" [('?', '.')],\n",
" [('Marke', 'NN')],\n",
" [('it', 'PRP')],\n",
" [('Horatio', 'NN')],\n",
" [('Hora', 'NNS')],\n",
" [('.', '.')],\n",
" [('Most', 'JJS')],\n",
" [('like', 'IN')],\n",
" [(':', ':')],\n",
```

```
" [('It', 'PRP')],\n",
" [('harrowes', 'NN')],\n",
" [('me', 'PRP')],\n",
" [('with', 'IN')],\n",
" [('fear', 'NN')],\n",
" [('&', 'CC')],\n",
" [('wonder', 'NN')],\n",
" [('Barn', 'NN')],\n",
" [('.', '.')],\n",
" [('It', 'PRP')],\n",
" [('would', 'MD')],\n",
" [('be', 'VB')],\n",
" [('spoke', 'NN')],\n",
" [('too', 'RB')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [('Question', 'NN')],\n",
" [('it', 'PRP')],\n",
" [('Horatio', 'NN')],\n",
" [('Hor', 'NN')],\n",
" [('.', '.')],\n",
" [('What', 'WP')],\n",
" [('art', 'NN')],\n",
" [('thou', 'NN')],\n",
" [('that', 'IN')],\n",
" [('vsurp', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('st', 'NN')],\n",
" [('this', 'DT')],\n",
" [('time', 'NN')],\n",
" [('of', 'IN')],\n",
" [('night', 'NN')],\n",
" [(',', ',')],\n",
" [('Together', 'RB')],\n",
" [('with', 'IN')],\n",
" [('that', 'IN')],\n",
" [('Faire', 'NN')],\n",
" [('and', 'CC')],\n",
" [('Warlike', 'IN')],\n",
" [('forme', 'NN')],\n",
" [('In', 'IN')],\n",
" [('which', 'WDT')],\n",
" [('the', 'DT')],\n",
" [('Maiesty', 'NN')],\n",
" [('of', 'IN')],\n",
" [('buried', 'VBN')],\n",
" [('Denmarke', 'NN')],\n",
" [('Did', 'NN')],\n",
" [('sometimes', 'RB')],\n",
" [('march', 'NN')],\n",
" [(':', ':')],\n",
" [('By', 'IN')],\n",
" [('Heauen', 'NN')],\n",
" [('I', 'PRP')],\n",
" [('charge', 'NN')],\n",
" [('thee', 'NN')],\n",
" [('speake', 'NN')],\n",
" [('Mar', 'NN')],\n",
```

```
" [('.', '.')],\n",
" [('It', 'PRP')],\n",
" [('is', 'VBZ')],\n",
" [('offended', 'VBN')],\n",
" [('Barn', 'NN')],\n",
" [('.', '.')],\n",
" [('See', 'VB')],\n",
" [(',', ',')],\n",
" [('it', 'PRP')],\n",
" [('stalkes', 'NNS')],\n",
" [('away', 'RB')],\n",
" [('Hor', 'NN')],\n",
" [('.', '.')],\n",
" [('Stay', 'NN')],\n",
" [(':', ':')],\n",
" [('speake', 'NN')],\n",
" [(';', ':')],\n",
" [('speake', 'NN')],\n",
" [(':', ':')],\n",
" [('I', 'PRP')],\n",
" [('Charge', 'NN')],\n",
" [('thee', 'NN')],\n",
" [(',', ',')],\n",
" [('speake', 'NN')],\n",
" [('.', '.')],\n",
" [('Exit', 'NN')],\n",
" [('the', 'DT')],\n",
" [('Ghost', 'NN')],\n",
" [('.', '.')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [(\"'\", \"''\")],\n",
" [('Tis', 'NN')],\n",
" [('gone', 'VBN')],\n",
" [(',', ',')],\n",
" [('and', 'CC')],\n",
" [('will', 'MD')],\n",
" [('not', 'RB')],\n",
" [('answer', 'NN')],\n",
" [('Barn', 'NN')],\n",
" [('.', '.')],\n",
" [('How', 'WRB')],\n",
" [('now', 'RB')],\n",
" [('Horatio', 'NN')],\n",
" [('?', '.')],\n",
" [('You', 'PRP')],\n",
" [('tremble', 'JJ')],\n",
" [('&', 'CC')],\n",
" [('look', 'NN')],\n",
" [('pale', 'NN')],\n",
" [(':', ':')],\n",
" [('Is', 'NN')],\n",
" [('not', 'RB')],\n",
" [('this', 'DT')],\n",
" [('something', 'NN')],\n",
" [('more', 'RBR')],\n",
" [('then', 'RB')],\n",
" [('Fantasie', 'NN')],\n",
```

```
" [('?', '.')],\n",
" [('What', 'WP')],\n",
" [('thinke', 'NN')],\n",
" [('you', 'PRP')],\n",
" [('on', 'IN')],\n",
" [(\"'\", \"''\")],\n",
" [('t', 'NN')],\n",
" [('?', '.')],\n",
" [('Hor', 'NN')],\n",
" [('.', '.')],\n",
" [('Before', 'IN')],\n",
" [('my', 'PRP$')],\n",
" [('God', 'NNP')],\n",
" [(',', ',')],\n",
" [('I', 'PRP')],\n",
" [('might', 'MD')],\n",
" [('not', 'RB')],\n",
" [('this', 'DT')],\n",
" [('beleeue', 'NN')],\n",
" [('Without', 'IN')],\n",
" [('the', 'DT')],\n",
" [('sensible', 'JJ')],\n",
" [('and', 'CC')],\n",
" [('true', 'JJ')],\n",
" [('auouch', 'JJ')],\n",
" [('Of', 'IN')],\n",
" [('mine', 'NN')],\n",
" [('owne', 'NN')],\n",
" [('eyes', 'NNS')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [('Is', 'NN')],\n",
" [('it', 'PRP')],\n",
" [('not', 'RB')],\n",
" [('like', 'IN')],\n",
" [('the', 'DT')],\n",
" [('King', 'VBG')],\n",
" [('?', '.')],\n",
" [('Hor', 'NN')],\n",
" [('.', '.')],\n",
" [('As', 'IN')],\n",
" [('thou', 'NN')],\n",
" [('art', 'NN')],\n",
" [('to', 'TO')],\n",
" [('thy', 'NN')],\n",
" [('selfe', 'NN')],\n",
" [(',', ',')],\n",
" [('Such', 'JJ')],\n",
" [('was', 'VBD')],\n",
" [('the', 'DT')],\n",
" [('very', 'RB')],\n",
" [('Armour', 'NN')],\n",
" [('he', 'PRP')],\n",
" [('had', 'VBD')],\n",
" [('on', 'IN')],\n",
" [(',', ',')],\n",
" [('When', 'WRB')],\n",
" [('th', 'NN')],\n",
```

```
" [(\"'\", \"''\")],\n",
" [('Ambitious', 'JJ')],\n",
" [('Norwey', 'NN')],\n",
" [('combatted', 'VBN')],\n",
" [(':', ':')],\n",
" [('So', 'RB')],\n",
" [('frown', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('d', 'NN')],\n",
" [('he', 'PRP')],\n",
" [('once', 'RB')],\n",
" [(',', ',')],\n",
" [('when', 'WRB')],\n",
" [('in', 'IN')],\n",
" [('an', 'DT')],\n",
" [('angry', 'JJ')],\n",
" [('parle', 'NN')],\n",
" [('He', 'PRP')],\n",
" [('smot', 'NN')],\n",
" [('the', 'DT')],\n",
" [('sledded', 'VBD')],\n",
" [('Pollax', 'NN')],\n",
" [('on', 'IN')],\n",
" [('the', 'DT')],\n",
" [('Ice', 'NN')],\n",
" [('.', '.')],\n",
" [(\"'\", \"''\")],\n",
" [('Tis', 'NN')],\n",
" [('strange', 'NN')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [('Thus', 'RB')],\n",
" [('twice', 'RB')],\n",
" [('before', 'IN')],\n",
" [(',', ',')],\n",
" [('and', 'CC')],\n",
" [('iust', 'NN')],\n",
" [('at', 'IN')],\n",
" [('this', 'DT')],\n",
" [('dead', 'JJ')],\n",
" [('houre', 'NN')],\n",
" [(',', ',')],\n",
" [('With', 'IN')],\n",
" [('Martiall', 'NN')],\n",
" [('stalke', 'NN')],\n",
" [(',', ',')],\n",
" [('hath', 'NN')],\n",
" [('he', 'PRP')],\n",
" [('gone', 'VBN')],\n",
" [('by', 'IN')],\n",
" [('our', 'PRP$')],\n",
" [('Watch', 'NN')],\n",
" [('Hor', 'NN')],\n",
" [('.', '.')],\n",
" [('In', 'IN')],\n",
" [('what', 'WP')],\n",
" [('particular', 'JJ')],\n",
" [('thought', 'NN')],\n",
```

```
" [('to', 'TO')],\n",
" [('work', 'NN')],\n",
" [(',', ',')],\n",
" [('I', 'PRP')],\n",
" [('know', 'VB')],\n",
" [('not', 'RB')],\n",
" [(':', ':')],\n",
" [('But', 'CC')],\n",
" [('in', 'IN')],\n",
" [('the', 'DT')],\n",
" [('grosse', 'NN')],\n",
" [('and', 'CC')],\n",
" [('scope', 'NN')],\n",
" [('of', 'IN')],\n",
" [('my', 'PRP$')],\n",
" [('Opinion', 'NN')],\n",
" [(',', ',')],\n",
" [('This', 'DT')],\n",
" [('boades', 'NNS')],\n",
" [('some', 'DT')],\n",
" [('strange', 'NN')],\n",
" [('erruption', 'NN')],\n",
" [('to', 'TO')],\n",
" [('our', 'PRP$')],\n",
" [('State', 'NN')],\n",
" [('Mar', 'NN')],\n",
" [('.', '.')],\n",
" [('Good', 'JJ')],\n",
" [('now', 'RB')],\n",
" [('sit', 'NN')],\n",
" [('downe', 'NN')],\n",
" [(',', ',')],\n",
" [('&', 'CC')],\n",
" [('tell', 'NN')],\n",
" [('me', 'PRP')],\n",
" [('he', 'PRP')],\n",
" [('that', 'IN')],\n",
" [('knowes', 'NN')],\n",
" [('Why', 'WRB')],\n",
" [('this', 'DT')],\n",
" [('same', 'JJ')],\n",
" [('strict', 'NN')],\n",
" [('and', 'CC')],\n",
" [('most', 'JJS')],\n",
" [('obseruant', 'NN')],\n",
" [('Watch', 'NN')],\n",
" [(',', ',')],\n",
" [('So', 'RB')],\n",
" [('nightly', 'RB')],\n",
" [('toyles', 'NNS')],\n",
" [('the', 'DT')],\n",
" [('subiect', 'NN')],\n",
" [('of', 'IN')],\n",
" [('the', 'DT')],\n",
" [('Land', 'NN')],\n",
" [(',', ',')],\n",
" [('And', 'CC')],\n",
" [('why', 'WRB')],\n",
```

```
" [('such', 'JJ')],\n",
" [('dayly', 'NN')],\n",
" [('Cast', 'NN')],\n",
" [('of', 'IN')],\n",
" [('Brazon', 'NN')],\n",
" [('Cannon', 'NN')],\n",
" [('And', 'CC')],\n",
" [('Forraigne', 'NN')],\n",
" [('Mart', 'NNP')],\n",
" [('for', 'IN')],\n",
" [('Implements', 'NNS')],\n",
" [('of', 'IN')],\n",
" [('warre', 'NN')],\n",
" [(':', ':')],\n",
" [('Why', 'WRB')],\n",
" [('such', 'JJ')],\n",
" [('impresse', 'NN')],\n",
" [('of', 'IN')],\n",
" [('Ship', 'NN')],\n",
" [('-', ':')],\n",
" [('wrights', 'NNS')],\n",
" [(',', ',')],\n",
" [('whose', 'WP$')],\n",
" [('sore', 'NN')],\n",
" [('Taske', 'NN')],\n",
" [('Do', 'VB')],\n",
" [(\"'\", \"''\")],\n",
" [('s', 'NN')],\n",
" [('not', 'RB')],\n",
" [('diuide', 'NN')],\n",
" [('the', 'DT')],\n",
" [('Sunday', 'NNP')],\n",
" [('from', 'IN')],\n",
" [('the', 'DT')],\n",
" [('weeke', 'NN')],\n",
" [(',', ',')],\n",
" [('What', 'WP')],\n",
" [('might', 'MD')],\n",
" [('be', 'VB')],\n",
" [('toward', 'IN')],\n",
" [(',', ',')],\n",
" [('that', 'IN')],\n",
" [('this', 'DT')],\n",
" [('sweaty', 'NN')],\n",
" [('hast', 'NN')],\n",
" [('Doth', 'NNP')],\n",
" [('make', 'VB')],\n",
" [('the', 'DT')],\n",
" [('Night', 'NN')],\n",
" [('ioynt', 'NN')],\n",
" [('-', ':')],\n",
" [('Labourer', 'NN')],\n",
" [('with', 'IN')],\n",
" [('the', 'DT')],\n",
" [('day', 'NN')],\n",
" [(':', ':')],\n",
" [('Who', 'WP')],\n",
" [('is', 'VBZ')],\n",
```

```
" [(\"'\", \"''\")],\n",
" [('t', 'NN')],\n",
" [('that', 'IN')],\n",
" [('can', 'MD')],\n",
" [('informe', 'NN')],\n",
" [('me', 'PRP')],\n",
" [('?', '.')],\n",
" [('Hor', 'NN')],\n",
" [('.', '.')],\n",
" [('That', 'DT')],\n",
" [('can', 'MD')],\n",
" [('I', 'PRP')],\n",
" [(',', ',')],\n",
" [('At', 'IN')],\n",
" [('least', 'JJS')],\n",
" [('the', 'DT')],\n",
" [('whisper', 'NN')],\n",
" [('goes', 'VBZ')],\n",
" [('so', 'RB')],\n",
" [(':', ':')],\n",
" [('Our', 'PRP$')],\n",
" [('last', 'JJ')],\n",
" [('King', 'VBG')],\n",
" [(',', ',')],\n",
" [('Whose', 'VB')],\n",
" [('Image', 'NN')],\n",
" [('euen', 'NN')],\n",
" [('but', 'CC')],\n",
" [('now', 'RB')],\n",
" [('appear', 'VB')],\n",
" [(\"'\", \"''\")],\n",
" [('d', 'NN')],\n",
" [('to', 'TO')],\n",
" [('vs', 'NN')],\n",
" [(',', ',')],\n",
" [('Was', 'NN')],\n",
" [('(', '(')],\n",
" [('as', 'IN')],\n",
" [('you', 'PRP')],\n",
" [('know', 'VB')],\n",
" [(')', ')')],\n",
" [('by', 'IN')],\n",
" [('Fortinbras', 'NNS')],\n",
" [('of', 'IN')],\n",
" [('Norway', 'RB')],\n",
" [(',', ',')],\n",
" [('(', '(')],\n",
" [('Thereto', 'NN')],\n",
" [('prick', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('d', 'NN')],\n",
" [('on', 'IN')],\n",
" [('by', 'IN')],\n",
" [('a', 'DT')],\n",
" [('most', 'JJS')],\n",
" [('emulate', 'NN')],\n",
" [('Pride', 'NN')],\n",
" [(')', ')')],\n",
```

```
" [('Dar', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('d', 'NN')],\n",
" [('to', 'TO')],\n",
" [('the', 'DT')],\n",
" [('Combate', 'NN')],\n",
" [('.', '.')],\n",
" [('In', 'IN')],\n",
" [('which', 'WDT')],\n",
" [(',', ',')],\n",
" [('our', 'PRP$')],\n",
" [('Valiant', 'NN')],\n",
" [('Hamlet', 'NN')],\n",
" [(',', ',')],\n",
" [('(', '(')],\n",
" [('For', 'IN')],\n",
" [('so', 'RB')],\n",
" [('this', 'DT')],\n",
" [('side', 'NN')],\n",
" [('of', 'IN')],\n",
" [('our', 'PRP$')],\n",
" [('knowne', 'NN')],\n",
" [('world', 'NN')],\n",
" [('esteem', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('d', 'NN')],\n",
" [('him', 'PRP')],\n",
" [(')', ')')],\n",
" [('Did', 'NN')],\n",
" [('slay', 'NN')],\n",
" [('this', 'DT')],\n",
" [('Fortinbras', 'NNS')],\n",
" [(':', ':')],\n",
" [('who', 'WP')],\n",
" [('by', 'IN')],\n",
" [('a', 'DT')],\n",
" [('Seal', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('d', 'NN')],\n",
" [('Compact', 'JJ')],\n",
" [(',', ',')],\n",
" [('Well', 'RB')],\n",
" [('ratified', 'VBN')],\n",
" [('by', 'IN')],\n",
" [('Law', 'NN')],\n",
" [(',', ',')],\n",
" [('and', 'CC')],\n",
" [('Heraldrie', 'NN')],\n",
" [(',', ',')],\n",
" [('Did', 'NN')],\n",
" [('forfeite', 'NN')],\n",
" [('(', '(')],\n",
" [('with', 'IN')],\n",
" [('his', 'PRP$')],\n",
" [('life', 'NN')],\n",
" [(')', ')')],\n",
" [('all', 'DT')],\n",
" [('those', 'DT')],\n",
```

```
" [('his', 'PRP$')],\n",
" [('Lands', 'NNS')],\n",
" [('Which', 'WDT')],\n",
" [('he', 'PRP')],\n",
" [('stood', 'NN')],\n",
" [('seiz', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('d', 'NN')],\n",
" [('on', 'IN')],\n",
" [(',', ',')],\n",
" [('to', 'TO')],\n",
" [('the', 'DT')],\n",
" [('Conqueror', 'NN')],\n",
" [(':', ':')],\n",
" [('Against', 'IN')],\n",
" [('the', 'DT')],\n",
" [('which', 'WDT')],\n",
" [(',', ',')],\n",
" [('a', 'DT')],\n",
" [('Moity', 'NN')],\n",
" [('competent', 'NN')],\n",
" [('Was', 'NN')],\n",
" [('gaged', 'VBN')],\n",
" [('by', 'IN')],\n",
" [('our', 'PRP$')],\n",
" [('King', 'VBG')],\n",
" [(':', ':')],\n",
" [('which', 'WDT')],\n",
" [('had', 'VBD')],\n",
" [('return', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('d', 'NN')],\n",
" [('To', 'TO')],\n",
" [('the', 'DT')],\n",
" [('Inheritance', 'NN')],\n",
" [('of', 'IN')],\n",
" [('Fortinbras', 'NNS')],\n",
" [(',', ',')],\n",
" [('Had', 'VBD')],\n",
" [('he', 'PRP')],\n",
" [('bin', 'NN')],\n",
" [('Vanquisher', 'NN')],\n",
" [(',', ',')],\n",
" [('as', 'IN')],\n",
" [('by', 'IN')],\n",
" [('the', 'DT')],\n",
" [('same', 'JJ')],\n",
" [('Cou', 'NN')],\n",
" [(\"'\", \"''\")],\n",
" [('nant', 'NN')],\n",
" [('And', 'CC')],\n",
" [('carriage', 'NN')],\n",
" [('of', 'IN')],\n",
" [('the', 'DT')],\n",
" [('Article', 'NN')],\n",
" [('designe', 'NN')],\n",
" [(',', ',')],\n",
" [('His', 'PRP$')],\n",
```

```
     " [('fell', 'VBD')],\n",
     " ...]"
    ]
   },
   "execution_count": 37,
   "metadata": {},
   "output_type": "execute_result"
  }
 ],
 "source": [
  "hamlet_pos"
 ]
},
{
 "cell_type": "code",
 "execution_count": 38,
 "id": "72f00bea",
 "metadata": {},
 "outputs": [],
 "source": [
  "hamlet_nnp = []\n",
  "for each_pos in hamlet_pos:\n",
  "    if each_pos[0][1] =='NNP':\n",
  "        hamlet_nnp.append(each_pos[0][0])"
 ]
},
{
 "cell_type": "code",
 "execution_count": 39,
 "id": "25318ece",
 "metadata": {},
 "outputs": [
  {
   "data": {
    "text/plain": [
     "['William',\n",
     " 'Francisco',\n",
     " 'Francisco',\n",
     " 'Westward',\n",
     " 'Bell',\n",
     " 'God',\n",
     " 'Mart',\n",
     " 'Sunday',\n",
     " 'Doth',\n",
     " 'Doth',\n",
     " 'God',\n",
     " 'Day',\n",
     " 'Sea',\n",
     " 'Wherein',\n",
     " 'Kingdome',\n",
     " 'Bedrid']"
    ]
   },
   "execution_count": 39,
   "metadata": {},
   "output_type": "execute_result"
  }
 ],
```

```
 "source": [
  "hamlet_nnp"
 ]
},
{
 "cell_type": "code",
 "execution_count": 41,
 "id": "623ae759",
 "metadata": {},
 "outputs": [],
 "source": [
  "same_names = [\"William\", \"william\", \"Reading\", \"reading\"]"
 ]
},
{
 "cell_type": "code",
 "execution_count": 42,
 "id": "7026b34b",
 "metadata": {},
 "outputs": [
  {
   "name": "stdout",
   "output_type": "stream",
   "text": [
    "[('William', 'NNP')]\n",
    "[('william', 'NN')]\n",
    "[('Reading', 'VBG')]\n",
    "[('reading', 'NN')]\n"
   ]
  }
 ],
 "source": [
  "for each_name in same_names:\n",
  "    print(nltk.pos_tag([each_name]))"
 ]
},
{
 "cell_type": "code",
 "execution_count": 43,
 "id": "4bbd1e26",
 "metadata": {},
 "outputs": [],
 "source": [
  "from nltk import ne_chunk"
 ]
},
{
 "cell_type": "code",
 "execution_count": 44,
 "id": "9958f01a",
 "metadata": {},
 "outputs": [],
 "source": [
  "NE_sent = \"The US Presidebt stays in the White House\""
 ]
},
{
 "cell_type": "code",
```

```
 "execution_count": 45,
 "id": "4683b9d6",
 "metadata": {},
 "outputs": [],
 "source": [
  "NE_tokens = word_tokenize(NE_sent)"
 ]
},
{
 "cell_type": "code",
 "execution_count": 46,
 "id": "c57964c0",
 "metadata": {},
 "outputs": [],
 "source": [
  "NE_tags = nltk.pos_tag(NE_tokens)"
 ]
},
{
 "cell_type": "code",
 "execution_count": 47,
 "id": "8e04e257",
 "metadata": {},
 "outputs": [],
 "source": [
  "NE_NER = ne_chunk(NE_tags)"
 ]
},
{
 "cell_type": "code",
 "execution_count": 48,
 "id": "8c270ee0",
 "metadata": {},
 "outputs": [
  {
   "name": "stdout",
   "output_type": "stream",
   "text": [
    "(S\n",
    "  The/DT\n",
    "  (ORGANIZATION US/NNP)\n",
    "  Presidebt/NNP\n",
    "  stays/VBZ\n",
    "  in/IN\n",
    "  the/DT\n",
    "  (FACILITY White/NNP House/NNP))\n"
   ]
  }
 ],
 "source": [
  "print(NE_NER)"
 ]
},
{
 "cell_type": "code",
 "execution_count": 50,
 "id": "6b1b6c30",
 "metadata": {},
```

```
 "outputs": [],
 "source": [
  "NE_sent2= \"The state of New York touches the Atlantic Ocean\""
 ]
},
{
 "cell_type": "code",
 "execution_count": 51,
 "id": "d278aba1",
 "metadata": {},
 "outputs": [
  {
   "name": "stdout",
   "output_type": "stream",
   "text": [
    "(S\n",
    "  The/DT\n",
    "  state/NN\n",
    "  of/IN\n",
    "  (GPE New/NNP York/NNP)\n",
    "  touches/VBZ\n",
    "  the/DT\n",
    "  (ORGANIZATION Atlantic/NNP Ocean/NNP))\n"
   ]
  }
 ],
 "source": [
  "print(ne_chunk(nltk.pos_tag(word_tokenize(NE_sent2))))"
 ]
},
{
 "cell_type": "code",
 "execution_count": 53,
 "id": "8ace5347",
 "metadata": {},
 "outputs": [],
 "source": [
  "NE_sent3 = \"Apple is a fruit and Apple is a Company's name\""
 ]
},
{
 "cell_type": "code",
 "execution_count": 54,
 "id": "fa8a241f",
 "metadata": {},
 "outputs": [
  {
   "name": "stdout",
   "output_type": "stream",
   "text": [
    "(S\n",
    "  (GPE Apple/NNP)\n",
    "  is/VBZ\n",
    "  a/DT\n",
    "  fruit/NN\n",
    "  and/CC\n",
    "  (PERSON Apple/NNP)\n",
    "  is/VBZ\n",
```

```
      "  a/DT\n",
      "  (ORGANIZATION Company/NN)\n",
      "  's/POS\n",
      "  name/NN)\n"
     ]
    }
   ],
   "source": [
    "print(ne_chunk(nltk.pos_tag(word_tokenize(NE_sent3))))"
   ]
  },
  {
   "cell_type": "code",
   "execution_count": null,
   "id": "22d5918b",
   "metadata": {},
   "outputs": [],
   "source": []
  },
  {
   "cell_type": "code",
   "execution_count": null,
   "id": "fb62afcf",
   "metadata": {},
   "outputs": [],
   "source": []
  }
 ],
 "metadata": {
  "kernelspec": {
   "display_name": "Python 3 (ipykernel)",
   "language": "python",
   "name": "python3"
  },
  "language_info": {
   "codemirror_mode": {
    "name": "ipython",
    "version": 3
   },
   "file_extension": ".py",
   "mimetype": "text/x-python",
   "name": "python",
   "nbconvert_exporter": "python",
   "pygments_lexer": "ipython3",
   "version": "3.11.5"
  }
 },
 "nbformat": 4,
 "nbformat_minor": 5
}
```