



Shakespeare

A Natural Language Processing Exploration





Why did I select
Shakespeare?

The Problems I Expected to Encounter

- Archaic Language vs. Modern Stop Word List
- Nuance and Innuendo
- Dialects and Other Languages

The Problems I Didn't Anticipate

- Scene Direction and Stage Directions
- Contractions and Elisions
- Processing Text
- Shakespeare's Stop Words are not our Stop Words



A Similarity Comparison Between Shakespeare's Plays

5 Topics

Topic 0 - Courtly Life

great, honour, noble, pray, life, poor, cannot, could, nature, nothing, look, fear, heaven, put, might

Topic 1 - Tragic Romance

love, night, sweet, look, madam, eyes, true, death, cannot, wit, word, nothing, dead, dear, eye

Topic 2 - Political Dramas

death, blood, life, hand, grace, crown, soul, fear, dead, live, heaven, die, noble, look, hast

Topic 3 - War

great, blood, love, sweet, true, hand, name, peace, many, faith, honour, every, look, war, full

Topic 4 - Marriage Comedies

pray, love, husband, house, woman, marry, life, better, name, signior, home, leave, money, look, till



Recurring Themes in all Shakespeare's Plays

The Grouped Plays



Henry VI Part 2
Titus Andronicus
Richard III
Henry VI Part 3

Titus Andronicus

Titus Andronicus

Richard III

Henry VI Part 3

The Grouped Plays



Play

As You Like It

Twelfth Night

Much Ado About Nothing



The Grouped Plays

Play

A Midsummer Nights Dream

Two Gentlemen Of Verona

Loves Labours Lost

Romeo And Juliet



The Grouped Plays

Play

Henry IV

Henry V

Troilus And Cressida



[illegible]

The Grouped Plays

Play

Henry VI Part 1

Richard II

King John

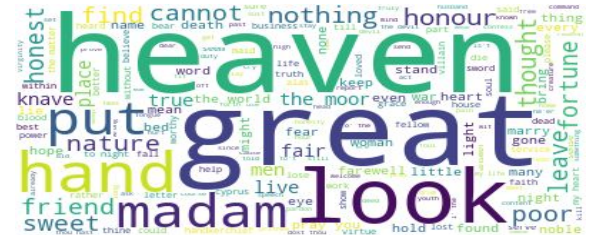


The Grouped Plays

Play

Alls Well That Ends Well

Othello



Resources:

pickle

pandas

string

re

numpy

nltk

sklearn

PIL

wordcloud



Additional Slides

Best Similarity For Each Play

A Comedy Of Errors is best matched with Merry Wives Of Windsor at 97.52% similarity.
A Midsummer Nights Dream is best matched with Two Gentlemen Of Verona at 98.14% similarity.
A Winters Tale is best matched with The Tempest at 99.69% similarity.
Alls Well That Ends Well is best matched with Othello at 96.66% similarity.
Antony And Cleopatra is best matched with Cymbeline at 97.26% similarity.
As You Like It is best matched with Twelfth Night at 97.69% similarity.
Coriolanus is best matched with Antony And Cleopatra at 87.51% similarity.
Cymbeline is best matched with The Tempest at 99.96% similarity.
Hamlet is best matched with King Lear at 95.73% similarity.
Henry IV is best matched with Henry V at 91.87% similarity.
Henry V is best matched with Troilus And Cressida at 93.37% similarity.
Henry VI Part 1 is best matched with Richard II at 99.27% similarity.
Henry VI Part 2 is best matched with Titus Andronicus at 98.86% similarity.
Henry VI Part 3 is best matched with Richard III at 98.11% similarity.
Henry VIII is best matched with A Winters Tale at 97.58% similarity.

Best Similarity For Each Play

Julius Caesar is best matched with Macbeth at 93.88% similarity.

King John is best matched with Richard II at 95.2% similarity.

King Lear is best matched with Cymbeline at 96.41% similarity.

Loves Labours Lost is best matched with A Midsummer Nights Dream at 90.73% similarity.

Measure For Measure is best matched with Merchant Of Venice at 84.28% similarity.

Merchant Of Venice is best matched with Taming Of The Shrew at 96.96% similarity.

Merry Wives Of Windsor is best matched with Taming Of The Shrew at 96.75% similarity.

Much Ado About Nothing is best matched with As You Like It at 95.06% similarity.

Othello is best matched with Alls Well That Ends Well at 96.66% similarity.

Pericles is best matched with Cymbeline at 95.67% similarity.

Richard II is best matched with Henry VI Part 1 at 99.27% similarity.

Richard III is best matched with Titus Andronicus at 99.14% similarity.

Romeo And Juliet is best matched with A Midsummer Nights Dream at 92.94% similarity.

Taming Of The Shrew is best matched with Merchant Of Venice at 96.96% similarity.

The Tempest is best matched with Cymbeline at 99.96% similarity.

Best Similarity For Each Play

Timon Of Athens is best matched with A Winters Tale at 99.55% similarity.

Titus Andronicus is best matched with Richard III at 99.14% similarity.

Troilus And Cressida is best matched with Henry V at 93.37% similarity.

Twelfth Night is best matched with As You Like It at 97.69% similarity.

Two Gentlemen Of Verona is best matched with A Midsummer Nights Dream at 98.14% similarity.

Macbeth is best matched with Antony And Cleopatra at 95.5% similarity.

Worst Similarity For Each Play

A Comedy Of Errors is worst matched with Loves Labours Lost at 0.0% similarity.
A Midsummer Nights Dream is worst matched with Henry VIII at 1.19% similarity.
A Winters Tale is worst matched with Henry V at 8.98% similarity.
Alls Well That Ends Well is worst matched with Richard II at 3.56% similarity.
Antony And Cleopatra is worst matched with Merry Wives Of Windsor at 0.68% similarity.
As You Like It is worst matched with Henry VI Part 2 at 7.2% similarity.
Coriolanus is worst matched with Merry Wives Of Windsor at 0.0% similarity.
Cymbeline is worst matched with Merry Wives Of Windsor at 2.62% similarity.
Hamlet is worst matched with Henry V at 13.87% similarity.
Henry Iv is worst matched with Two Gentlemen Of Verona at 2.68% similarity.
Henry V is worst matched with Two Gentlemen Of Verona at 0.61% similarity.
Henry VI Part 1 is worst matched with Merry Wives Of Windsor at 0.0% similarity.
Henry VI Part 2 is worst matched with Two Gentlemen Of Verona at 0.54% similarity.
Henry VI Part 3 is worst matched with Merry Wives Of Windsor at 1.83% similarity.
Henry VIII is worst matched with Loves Labours Lost at 0.0% similarity.

Worst Similarity For Each Play

Julius Caesar is worst matched with Merry Wives Of Windsor at 1.59% similarity.

King John is worst matched with Merry Wives Of Windsor at 0.37% similarity.

King Lear is worst matched with Henry V at 3.35% similarity.

Loves Labours Lost is worst matched with Henry VIII at 0.0% similarity.

Measure For Measure is worst matched with Loves Labours Lost at 2.06% similarity.

Merchant Of Venice is worst matched with Troilus And Cressida at 9.61% similarity.

Merry Wives Of Windsor is worst matched with Coriolanus at 0.0% similarity.

Much Ado About Nothing is worst matched with Henry VI Part 2 at 9.04% similarity.

Othello is worst matched with Richard II at 3.2% similarity.

Pericles is worst matched with Henry V at 9.84% similarity.

Richard II is worst matched with Merry Wives Of Windsor at 0.11% similarity.

Richard III is worst matched with Merry Wives Of Windsor at 2.76% similarity.

Romeo And Juliet is worst matched with Coriolanus at 0.0% similarity.

Taming Of The Shrew is worst matched with Coriolanus at 4.77% similarity.

The Tempest is worst matched with Merry Wives Of Windsor at 3.71% similarity.

Worst Similarity For Each Play

Timon Of Athens is worst matched with Henry VI Part 3 at 4.85% similarity.

Titus Andronicus is worst matched with Troilus And Cressida at 3.95% similarity.

Troilus And Cressida is worst matched with Merry Wives Of Windsor at 0.61% similarity.

Twelfth Night is worst matched with Henry VI Part 2 at 9.78% similarity.

Two Gentlemen Of Verona is worst matched with Coriolanus at 0.0% similarity.

Macbeth is worst matched with Merry Wives Of Windsor at 0.42% similarity.

Merry Wives of Windsor

Topic 0

wife, well, cuckold, fie, jealousy, wrong, hour, hearts, trust, names

Topic 1

love, say, let, sympathy, ha, see, young, john, reason, soldier

Topic 2

good, john, know, gar, heart, pray, well, tell, de, make

Topic 3

like, think, man, knave, clothes, basket, went, hot, woman, foul

Topic 4

let, away, pinch, fairies, one, time, way, marry, green, white