Week: 07. Module: B – Shakespeare on Page, Stage, Media

Unit: 15: Digital Shakespeare

Dr. Joseph Koyippally

Essentially Shakespeare has become digital. The integration of computer science into Shakespeare, or digital innovation of Shakespeare, had its forerunners. Much before the invention of the computer, Shakespeare scholars have tried different technological experimentations to address the complexity of the Shakespeare archive. Shakespeare editors have also used several experimentations before the digital intervention.

In this unit, we will be discussing 1) Digital Shakespeare, 2) Variant Shakespeare, 3) Digital Shakespeare production, 4) MIT Shakespeare Corpus, and 5) Archiving Shakespeare.

Digital Shakespeare

Even as computer has changed our approach to Shakespeare, his prominence in the digital terrain recognizes his cultural prominence also. With digital text, is a virtual text and we might wonder what digital approach to him does to Shakespeare. His centrality in digital humanities canon shows his importance in education and arts. However, no digital project can offer an unmediated view of his past or present. This makes a recognition of his presence across cultures important as it recognizes both his presence across cultures and also how different cultures have contributed to Shakespeare.

Till a few years ago digital Shakespeare only meant a searchable digitised version of his texts and databases. The digital archives disseminate rare and new material to a worldwide audience. Digital tools facilitate cross-cultural, collaborative, and interdisciplinary research on Shakespeare's texts and performances.

Besides plain text versions, now it also helps us to see virtual images of his Quarto and Folio editions; pictures, paintings, illustrations, audios and videos of his texts, criticism, and online discussion forums. We could search, see, download, store, and reproduce these digital Shakespeare data. Data-mining techniques also have become increasingly sophisticated helping one to place Shakespeare within the cyber infrastructure.

As his words are the best means to understand Shakespeare, a corpus analysis can illustrate the peculiar ways in which he has used them. By looking at the concentration of certain words and phrases in Shakespeare, one might unravel the working of Shakespeare's mind. It can throw light upon his sources, the influences upon him, and on his society. 'Text analysis' is an umbrella term. The algorithms developed by computer scientists, statisticians, and linguists help us to computationally analyse a text for word frequencies, co-occurrence, and statistically generate 'topics' to perform 'distant reading'.

The British Library's Shakespeare Quarto's Project and MIT's Shakespeare Performances in Asia, different "Internet editions" of Shakespeare, and the various electronic archives we have seen in Shakespeare and the Internet are major initiatives in bringing the Bard to the cyber stage. The Works of Donaldson, Holland and Onoeato, and Muller in *Shakespeare* 4 (2008) have looked at these issues critically.

Today, educational institutions, libraries, and software developers collaborate in creating digital applications of literary texts. This includes several platforms like text performance history, criticism, and media. The new media makes digital Shakespeare remain so prominent with his rhizomatic

presence in the digital sphere that as material objects and networks become models for post-human relations. This makes contemporary Shakespeare increasingly identified with digital Shakespeare. Recent studies (Donaldson; Galey; Lanier; Galey and Siemens; Galey; Muller; O'Neill, Shakespeare and Social Media) show the trend in digital Shakespeare initiatives. Shakespeare's digital recordings and broadcasts are already available and these are likely to lead to theories of digital dramaturgy, digital performances, and digitally enhanced Shakespeare. The digital versions, technology, and connectivity are likely to lead to digital production, reception, and research.

Even by understanding, for example, that the most frequently used content words in Shakespeare are 'lord' (3260), 'sir' (2972), 'king' (2965), 'good' (2855), 'come' (2537), 'let' (2083), 'love' (1926), 'man' (1846), 'like' (1805), 'say' (1671), 'know' (1662), 'make' (1629), 'speak' (1171), 'duke' (1095), 'tell' (1069), 'time' (1043), 'think' (1020) etc. could tell us about the social orientation of the Elizabethan society.

The difference in the use of "sister" (180) and "brother" (); different forms of forms "mother" (about 450) and of "father" (about 1000) tells us its gender politics. The frequent use of the first-person reflexive pronoun "myself" (26) in *Richard II*, and its increase in the king's dialogues, offer "impressive array of alternative ways of searching and viewing the text" (Best 155).

Another way to represent the Causes of death in Shakespeare Plays, given in Morningside College website tells us about the life in Shakespearean society

The concentration of such words in a particular period in, or a particular play, or even in a particular character can easily be analysed digitally. Specialised software like *Wordle* graphically represents word frequency. The web-based open-source text-analysis tool *Voyant* reads and interprets corpus. *Natural Language Toolkit* provides easy-to-use interfaces to over 50 corpora and lexical resources allow one to visualise linguistic data in different ways ranging from simple analysis to sophisticated ones. Some other open-source text analysis tools are Carrot2, Gensim, Orange, Stanbol etc.

The Folger Digital Texts freely gives high-quality digital texts of Shakespeare's plays, sonnets, and poems, to read online, download in five formats: XML, HTML, PDF, DOC (including or not including line numbers), TXT, and TEI Simple along with full source code in. The source code can help researchers and developers to develop apps and other projects compatible with its text, which follows the page numbering and layout of the Folger printed editions.

His digitisation has helped to place Shakespeare across the barriers of high culture, nationalities and languages and democratise the learning of Shakespeare. Social media plays a major role in this. Shakespeare has many Facebook pages, Twitter accounts, and YouTube videos. But, it is a paradox that although the Internet globalises the local, individual insularity increases among Shakespeare on the Internet as users who use the same the social media network and apps differ in deciding the information they want to receive.

Variant Shakespeare

The Globe edition (1864) of Shakespeare came out in the heydays of monoculturalism if not cultural assimilation. The book looked like the King James Bible in format and compactness. The logo on its title page was a blank globe marked with longitudes encircled by the Shakespearean motto: "One touch of nature makes the whole world kin". It appears to have asserted two things: that the Prime Meridien passes through Greenwich as was established in 1851, and the name Globe, though sounds similar to Shakespeare's theatre, actually refers to the world then ruled by the English Empire. English language, Shakespeare and the Bible seemed to assert the universality and globalised Englishness. It projected

that the Globe edition seemed to declare 'one Shakespeare for the entire world' suppressing any evidence of the editorial efforts of Cambridge University.

On the contrary, in post-imperialist days of multi-, if not pluricultural days, Shakespeare appears many. When Moby Shakespeare, the first Internet Shakespeare appeared, it was supposed to be a digital version of the Globe text. However, its agenda varied from that of the Globe Shakespeare. A digital text of Shakespeare means developing codes designed to store and transport data using eXtensible Markup Language (XML) or Hypertext Markup Language (HTML), which is not English as such. A digital Shakespeare compares different versions of Shakespeare and everything becomes valid, unlike the monolithic Globe edition. Longitude might still notionally pass through Greenwich, but Shakespeare everywhere is multicultural. Anyone involved in this naturally becomes conscious of Shakespeare's past editions and editors, and any modern netizen becomes aware of Shakespeare's multitudes.

The Shakespearean text, as the prefaces to his editions, remind us, manuscripts of none of his plays have survived. It is editors that have given us what we call Shakespeare today, and that makes Shakespeare a construct. None of his modern editions is purely based on any manuscript. They are not even based on the first printed versions of his works; rather, they are collated and corrected versions of his works from different sources and editorial traditions. The Norton Facsimile of the Shakespeare's First Folio, for example, is also a collated version of the different surviving copies of the First Folio, and therefore, a virtual text.

Right from the days of the First Folio editors Hemminge and Condell, editors have been grappling with the problem of Shakespeare's original text. Ever since Edmund Malone, the first modern editor of Shakespeare, compared the 'source texts' of the Folios with the variant Quartos, editors difficulty only got compounding difficult. When first Variorum editions were prepared, it became acute. Modern Variorum editions are born out of Isaac Reed's Variorum editions of 1803 and 1813 which he based on all the extant variant texts, and James Boswell's more meticulous Third Variorum edition (1821). These led to the production of the modern variorum edition of Shakespeare by HH Furnass (1870). The problem they encountered was how to mark up the differences.

British library's Webpage 'Treasures in Full Shakespeare in Quarto' compares the texts of Shakespeare's plays. https://www.bl.uk/treasures/SiqDiscovery/ui/search.aspx

An early example is Teena Rochfort Smith's *A Four-Text Edition of Shakespeare's Hamlet* (1883). It is a visually complex edition illustrating the differences between *Hamlet* texts. In four columns, she gives the first quarto (1603), the second quarto (1604/5), the first folio (1623) texts and her version. More than the layout, it is her use of the possibilities of Victorian typography that distinguishes it. She uses Roman, italic, Clarendon, sans serif, gothic, fonts and italic, bold and small capital formats and diacritical marks to visually mark up the orthographic variants (Thompson). Her visual markup typography could be compared to the tags coders use to embed information in digital texts.

A four-text edition of Shakespeare's Hamlet: 1. quarto 1, 1603 -- 2. quarto 2, 1604 -- 3. folio 1, 1623 -- 4. a revised text: in parallel columns / edited by Teena Rochfort Smith. From Folger Digital Image Collection. Digital Image File Name: 6796)

The cultural implication of Shakespeare's mechanical reproduction was a major concern in the nineteenth century. What happens to Shakespeare when he is converted into a machine-readable format with the rise of electronic computing bothered many? During World War II, durability and mobility of texts became a major concern. It was with the invention of cryptographic approaches to

Shakespeare's texts became prominent, with Charlton Hinman and Fredson Bowers leading the move. The transformation of Shakespeare's texts into data became ever since texts were converted into data and information.

There are serious discussions about a Shakespeare Folio known as the Pavier Quartos or the 'False Folio' of 1619, which came four years before the publication of the First Folio (1623). In this first attempt to collect Shakespeare's works into a single volume, William Jaggard printed ten plays—*Henry V, King Lear, The Merchant of Venice, The Merry Wives of Windsor, A Midsummer Night's Dream, Pericles, Prince of Tyre, Sir John Oldcastle, A Yorkshire Tragedy, Henry VI Part 3 and Henry VI Part 3*—attributed to Shakespeare.

Shakespeare's text is not a stable linguistic structure; and so, the virtual text created by the digital media is only as challengeable as his 'real' texts we come across. This might lead to an interesting question. With advanced computing programming, questions like who will be the next editor of Shakespeare, what his texts will be like in the coming years, and will they be read, studied or performed are increasingly becoming relevant.

Digital Shakespeare Production

As we have seen earlier, digital Shakespeare helps you to create new texts. For example, you can compare the Frist Quarto of Hamlet (1603) with the Fourth Quarto (1622) at the British Library, and create a compound text of Hamlet out of two quartos. But it is not enough. Do we need to trust the past scholarship or do we need to take a fresh look at all the different quarto and Folio editions of Hamlet before we create our text of *Hamlet*?

A traditional teacher is bound to fail in convincing his students that Shakespeare's printed text is more important than his performed text at a time when the author is dead and not he, but his editors get the copyright for their virtual editions of Shakespeare. You might willingly believe that the author is dead thanks to their theory classes and prefer to watch Shakespeare on the Internet and live in a digital world which foreground intertextuality thanks to hypertext links. When text gets transformed into other texts, pictures, or videos of performances from different cultures and times is not fixed.

Digitization of Shakespeare developed even before 1976, when machine-readable formats (ASCII, TXT, HTML etc.), and mark-up languages (COCOA, SGML, HTML, XML), etc. developed. digital Shakespeare marked by interoperability and cross-referencing created a virtual text by comparing different versions of texts; modernising or retaining his spellings; adding or deleting texts, using tools and methods for organizing, storing and retrieving data. Most of the websites hosting Shakespeare supplement his text by giving explanations, comments, and additional information. The first web Shakespeare, MIT's *The Complete Works of William Shakespeare*, continue to give plain Shakespeare text. However, resources like Shakespeare Navigators, gives notes scene summaries, annotations, and key critical texts on the major tragedies; and Open Source Shakespeare gives concordance, lists of characters and lines.

Open Source Shakespeare is an experiment in literary technology. It discusses the history of open source Shakespeare, the history of the globe Shakespeare and its characteristics, the Moby Shakespeare revolution, editing and structure of open-source Shakespeare, displaying the texts, and the future of open source Shakespeare. It also gives in three appendixes its database structure and documentation, marked-up play text, and parser source code.

Digital Shakespeare became a major issue when Hypertext, the software system allowing extensive cross-referencing, revolutionised the Internet since 1994 and became key to interoperability between programming languages. Grady Ward's public-domain *Moby Shakespeare* or the *Complete Works of William Shakespeare* revolutionised digital Shakespeare in the domain of the interoperable open network. It helped people across the world to access, encode, transcode, and redistribute Shakespeare files in the digital environment.

Editorial techniques, spelling, and variant readings of texts continue to be major issues in creating the internet, Shakespeare. The Riverside Shakespeare's modern-spelling *Complete Works* by Marvin Spevack (1969), Trevor Howard-Hill's old spelling Shakespeare concordance (1969) are popular Shakespeare digital texts. The old spelling in the facsimile of the First Folio by Sidney Lee (1902) and Jonathan Bate's and Eric Rasmussen's modern spelling versions of the First Folio are also popular Shakespeare resource. Grady Ward's Moby Shakespeare's (1995) with its modern, regularized spelling Globe text (1864) was the most popular version on the Internet.

An issue with digital Shakespeare is its conflict with traditional reference system causing problems to cite, quote or concord verses, lines or words. Digital reference system (DRS) is being but needs an overhauling of the traditional method. The traditional method is using short titles (e.g. *Oth., Ham., Rom*. etc.) followed by act and scene and line numbers as in *Oth*. I.ii.15-19. these abbreviations are based on the First Folio titles. While some two-letter abbreviations *like JC, TN, WT* remain unchanged since 1805; Some three-letter abbreviations (*LLL, MND, Tmp*); for titles with personal names (*COR, CYM, HAM, MAC, OTH, TIT*); adding lower-case for function words (*AaC, CoE, MfM, MoV, RaJ, TaC, ToA, ToS*), four-letter abbreviations (*TGoV, MWoW, MAaN, AyLi, AwEw*); and for lay with the names of kings (*1KH4, 2KH4, KH5, 1KH6, 2KH6, 3KH6, KR3, KH8, KJN, KLR, KR2*) are proposed. However, it is a long way to go in terms of standardization.

Computational coding shows the inadequacies of the traditional Act-scene-line division. Even in the First Folio itself, it is irregular. Six plays have no act-scene division, ten have only act division, and 20 plays have both act and scene division. The later editors found it difficult and Nicholas Rowe (1709) divides *Titus Andronicus* into 7 scenes while Thomas Hanmer (1750) divides it into 34 scenes, and the 'ideal' text of WG Clark and WA Wright (1860/6) which conflated quarto and folio texts has 14 scenes.

Whether to add the para-textual elements like textual divisions, character headings, stage directions, etc. as found in the standard editions affect digital Shakespeare, posing line numbering system is an issue. Stanley Wells and Gary Taylor debates with digital editors like Lou Burnard, John Lavagnino on Sequential Line Numbering (SLN), Through Line Numbering (TLN), Fixed Line Numbering (FLN), and Key Line Numbering (KLN). Charlton Hinman's Through Line Numbering TLN) in *The Norton Facsimile of the First Folio of Shakespeare* (1968), his 'ideal' but virtual First Folio edition produced by collating the good pages from different Folios uses unique TLN. It counts only the speech prefixes and para-textual elements like stage directions but provides for cross-referencing between TLN and SLN allowing linear and sequential approaches.

Digital Shakespeare is not limited to textual images. His digital enhanced productions hold much more for us in the future. The Commonwealth Shakespeare Company's 2019 production of *Hamlet* using virtual reality gives you a different experience with a virtual reality headgear. It allows you to sit at home to watch *Hamlet 360: Thy Father's Spirit*, and it puts you at the very centre of the tragedy, as the Ghost. Commonwealth Shakespeare Company, the Virtual Reality Company Sensorium and Google collaborated to create this one hour play.

Using the VR device, you can watch the events as they unfold. You can also turn your head at 360 degrees to watch although you cannot move, and become part of the play. It uses VR possibilities in many scenes an example is the "To be or not to be" scene, delivered by a Hamlet in the Bathtub. Here, you feel that you, as the Ghost, are sitting on Hamlet's chest as he is underwater and you hear the soliloquy. Only in the very last scene, Ghost moves away, after walking around the dead bodies.

MIT Shakespeare Corpus

Many websites help us statistically analyse Shakespeare. For example, Open source Shakespeare gives us Shakespeare statistics as 884,421 words in 43 works, with 28,829 unique word forms (43.3% of total word forms) of which 12,493 occur only once. His plays contain 34,895 speeches by 1,223 characters.

Let us look at some of the possibilities of applying Digital Humanities methods to on MIT online Shakespeare (shakespeare.mit.edu). If you analyse it by using 4.0 Voyant Tools l, (https://voyant-tools.org, Shakespeare's corpus has 37 plays, 880,555 words, 25,063 unique word forms, and a vocabulary density of 0.028, and 16.0 average words per sentence.

The longest document in terms of words is *Hamlet* (32396 words), and is followed by Richard III (31846); *Coriolanus* (29485); *Cymbeline* (29204); *Henry IV Part 1* (28225); *Othello* (28059); and *King Lear* (27986). The shortest in terms of words is *The Comedy of Errors* (16339)and is followed by *A Midsummer Night's Dream* (17332); The Tempest (17582); *Macbeth* (18414); *Two Gentlemen of Verona* (18443); Pericles (19827); *Timon of Athens* (19833); Julius Caesar (21033).

In terms of vocabulary density, the highest is *The Tempest* with an average density of 0.182. It is followed by *Macbeth* (0.182); *A Midsummer Night's Dream* (0.173); *Timon of Athens*. (0.168); *Henry VI Part 1* (0.167); and *Pericles* (0.167). The Romances have a high density of words among Shakespeare plays. The lowest vocabulary density is in *Richard III* which has a low density of 0.129. It is followed by *Much Ado About Nothing* (0.131); *Othello* (0.135); *Merry Wives of Windsor* (0.136); *Henry VI Part 1* (0.137); *Julius Caesar* (0.137); *Coriolanus* (0.138)

If one looks at the Average Words per sentence, the highest is King Richard II with an average of 21.5 words per sentence. *King John* (20.4); *Henry V* (20.3); *Henry VIII* (19.4); *Titus Andronicus* (19.1); T*he Winter's Tale* (18.5); *The Merchant of Venice* (18.0) follow. The lowest Words per sentence is in Othello (13.6); *Merry Wives of Windsor* (13.6); *Two Gentlemen of Verona* (13.9); *King Lear* (14.2); *Troilus and Cressida* (14.3); *Loves Labours Lost* (14.3); *Antony and Cleopatra* (14.3); *Timon of Athens* (14.5); *Julius Caesar* (14.7);

Without excluding function words, the most frequently used fifty words in Shakespeare corpus are: the (27961), and (25708), I (20576), to (19437), of (17135), a (14625), you (13712), my (12195), in (10770), that (10655), is (9056), not (8412), me (7673), it (7667), with (7614), for (7495), be (6818), his (6696), your (6687), this (6565), he (6289), but (6185), have (5891), as (5681), thou (5321), him (5198), so (4956), will (4931), what (4463), all (3850), her (3835), thy (3772), no (3752), do (3720), by (3693), shall (3564), if (3497), are (3400), we (3359), lord (3260), our (3112), thee (3065), on (3000), sir (2972), king (2965), good (2855), now (2801), o (2604), from (2602), and come (2537). The most frequently used content words in the corpus are: lord (3260), sir (2972), king (2965), good (2855), come (2537), enter (2388), let (2083), love (1926), man (1846), like (1805), say (1671), know (1662), make (1629), speak (1171), henry (1107), duke (1095), tell (1069), exeunt (1050), time (1043), think (1020), exit (972), heart (964), queen (957), lady (941), great (912), day (905), hear (878), men (871), death (864), away (845), father (845), life (842), hand (828), night (818), look (815), god (786), master (785), mistress (784), scene (782), true (779), ay (766), pray (747), sweet (742), fair (733), prince (693), Gloucester (687), old (660), blood (658), leave (651), son (646).

The most frequently used words are: shall (3594); lord (3346); king (3309); sir (3031); good (2883); come (2569); enter (2403); let (2108); love (1956); hath (1913); man (1890); like (1845); i'll (1778); say (1686); know (1670); make (1633); tis (1411); henry (1310); speak (1179); duke (1167); tell (1077); exeunt (1061); time (1048); think (1033); queen (991); lady (978); exit (976); heart (973); great (918); day (912); hear (886); men (884); art (880); act (877); death (872); doth (866); father (858); away (853); life (843); night (843); hand (840); scene (824); look (820); god (803); master (798); mistress (795); true (785); ay (766); pray (759); prince (757).

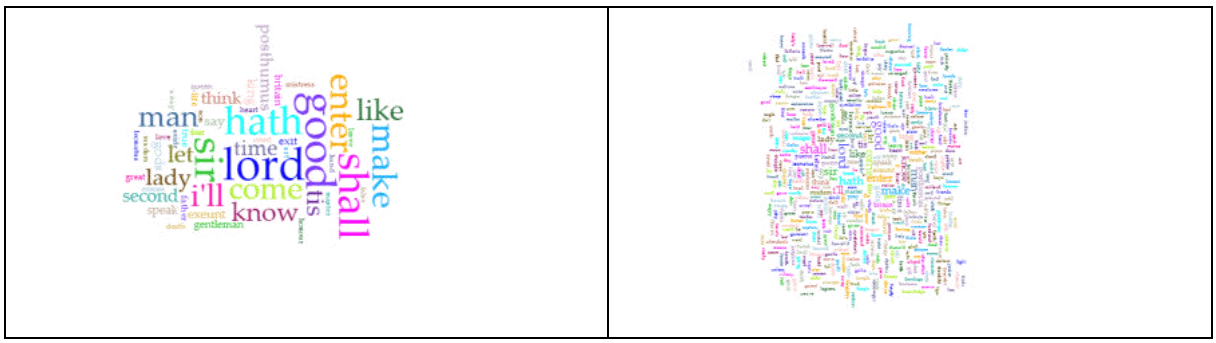
Using the following chronology –1590 (LLL), 1591 (CE, 6 Hen.1, Rom, TGV), 1592 (Tit), 1593 (6 Hen. 2, Rich. III0, 1594 (MV, MSND, Shr), 1595 (6 Hen.3, 1596 (KJ), 1597 (4 Hen. 1), 1597 (4 Hen. IV2, Rich. II), 1598 (AWW, AYLI), 1599 (Ham, JC, 5 Hen, MWW, TN, 1600 (MAAN, 1602 Tro, 1604 (MM, Oth), 1605 (Lear, 1605 Mac), 1607 (Ant, Tim), 1608 (Cor, Per), 1610 (Cym, WT), 1611 (Tmp), 1613 (8 Hen) one gets interesting results

A textual analysis gives the following data. The resultant Shakespeare corpus has 880,555 words with 25,063 unique word forms, vocabulary density of 0.028 and 16.0 average words per sentence. This study compares the individual Romance plays against Shakespearean corpus.

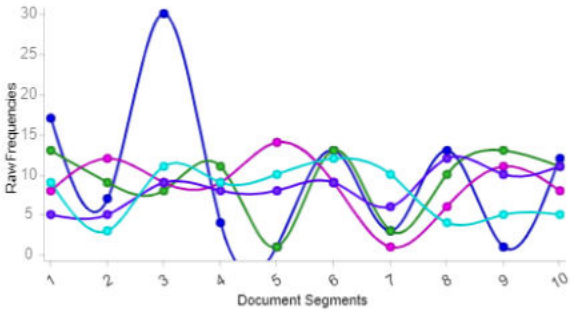
Text	Words	Unique words	Vocab. Density	AW/S
Cymbeline	28811	4280	0.153	16.1
Pericles	19449	3302	0.175	15.8
The Tempest	17344	3197	0.191	14.7
Winter's Tale	25806	3,907	0.154	18.3
Romances	91410	8,372	0.094	16.3
Shakespeare	880555	25,063	0.028	16.0

By analysing individual plays, we can graphically represent the content words in Shakespeare's Romances as follows. The most frequent words in Cymbeline are: lord (101); sir (92); shall (87); good (83); hath (78); enter (70); i'll (68); make (61); like (60); come (55); tis (55); man (53); let (52); know (48); lady (47); time (45); think (40); second (39); posthumous (38); gods (35); king (35); say (35); exeunt (34); exit (34); Britain (33); speak (33); father (32); gentleman (32); life (32); true (32); art (31); fear (31); great (31); aside (30); honour (29); love (29); heart (28); leave (28); son (28); hand (27); master (27); mistress (27); queen (27); roman (27); away (26); court (26); death (26); false (26); leonatus (26); madam (25); pisanio (25); mother (24); poor (24); pray (24); dead (23); imogen (23); best (22); brother (22); die (22)

Content worlds in Cymbeline may be represented in a simple manner or more complex manner as follows:



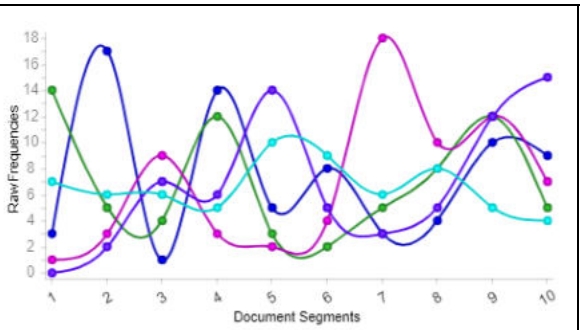
Considering that the top five words [lord (101); sir (92); shall (87); good (83); hath (78)] are used as follows in the ten segments of the text:



Similarly, the top colocation of the word 'sir' could be seen as follows:

The same concerning *Pericles* which has 18,913 words, 3,302 unique word forms, Vocabulary Density of 0.175, and Average Words Per Sentence: 15.8 is as follows:

Most frequent words in the corpus: lord (74); like (70); come (69); sir (69); enter (66); shall (64); good (56); bawd (49); gods (46); i'll (46); make (45); king (42); daughter (41); sea (37); second (37); hath (35); fisherman (34); know (34); pericles (34); men (32); tis (32); say (30); man (29); exeunt (28); tyre (28); gentleman (27); life (27); love (26); exit (25); let (25); tell (25); honour (24); dead (23); father (23); marina (23); prince (23); think (23); eyes (22); knight (22); tarsus (22); till (22); great (21); hear (20); doth (19); heaven (19); knights (19); look (19); place (19); fair (18); lady (18); speak (18); child (17); death (17); house (17); leave (17)



Top 5 words visualisation: lord (74); like (70); come (69); sir (69); enter (66);

The same concerning The Tempest which has 25,306 words, 3,907 unique word forms, Vocabulary Density of 0.154 and Average Words Per Sentence of 18.3 is as follows

Most frequent words in the corpus: sir (113); shall (111); good (104); come (87); lord (86); say (73); clown (72); i'll (65); like (57); shepherd (56); know (55); let (53); man (52); gentleman (50); king (49); tis (49); make (45); queen (44); father (43); think (43); camillo (41); enter (41); hath (41); life (40); speak (36); son (35); time (35); daughter (31); honour (31); lady (29); true (29); art (28); great (28); look (28); hand (27); bear (26); bohemia (26); heart (26); love (25); best (24); boy (24); business (24); prince (24); servant (24); better (23); death (23); hear (23); poor (23); thought (23); pray (22); court (21); fear (21); way (21); wife (21); brother (20); exeunt (20); exit (20); gone (20); hast (20)

The graphic representation of its most frequently used words—appear as

sir (113); shall (111); good (104); come (87); lord (86)	Collocations of the word 'sir'

The same concerning all the four Romances is 88,849 words, 8,372 unique word forms, vocabulary density of 0.094 and average words per sentence of 16.3 is as follows

Most frequent words in the corpus: sir (334); shall (323); good (287); lord (287); come (252); i'll (237); like (233); enter (221); make (192); hath (180); say (169); man (165); know (160); king (157); let (156); tis (155); time (119); think (118); father (117); life (114); gentleman (109); speak (107); art (104); daughter (104); lady (98); queen (98); exeunt (96); exit (93); great (92); honour (92); love (92); gods (91); second (90); son (88); true (87); sea (85); aside (84); hear (84); look (82); master (80); men (79); tell (78); heart (75); pray (75); best (74); poor (74); fear (73); clown (72); death (72); eyes (72); hand (70); leave (70); hast (69); dead (68); day (67); way (67); away (65); bring (64); mistress (63)

Archiving Shakespeare

Shakespeare has become fundamentally digital. This digital turn has its impact on our understanding of Shakespeare and his cultural afterlives. He has become the product of human-digital interaction and digital technologies, resources and cultures, directly and indirectly, influence the way we read, watch, research, and teach him. From the perspective of digital humanities, this turn is irreversible.

The digital turn has made his online texts readily available—texts here mean scanned copies, plain text formats, and performance texts—in various languages and cultures across times. Early English Books Online which makes available more than 125,000 titles in digital facsimile form and allows cross-searching with Early European Books also allows one to compare the variations in the early quarto editions of Shakespeare's plays with the First Folio.

In India also, online journals, newspapers, dissertations and other databases have changed the way research is carried out. The MHRD digital initiative of *INFLIBNET* which provides access to online journals, *Shodhganga*, which gives access to PhD thesis awarded by Indian universities, helps one to discover, access and manage rich, and diverse sources and data.

Every time-saving procedure also creates opportunities to waste time. Just like the new media like email or mobile phones also have their negative effects, digital Shakespeare also has its drawbacks. A reliable digital book costs more than a printed book although the former has the advantage of multiple simultaneous circulations.

Some important Shakespeare archives can be found at Folger Shakespeare Library, Washington DC, Shakespeare Centre Library & Archive, Stratford-upon-Avon, Harvard Theatre Collection, Harvard, Royal Shakespeare Company, Stratford-upon-Avon;

New York Public Library for the Performing Arts, New York; Performance Bibliographies at Shakespeare's Staging; National Theatre Archive, London; Horace Howard Furness Memorial Library, University of Pennsylvania, New York University's Shakespeare Studies page also leads one to many sites on Shakespeare.

Image collections from The Cleveland Press Shakespeare Photographs; Touchstone Project by the British Council Library

The internet sources like Open Source Shakespeare, Open Shakespeare, PlayShakespeare, Shakespeare Documented etc. provide very valuable resources.

MIT's Shakespeare Electronic Archive is easy to use the system. It provides links to electronic texts of Shakespeare's plays and the digital copies the primary materials. The archive gives access to the Shakespeare's Oxford Electronic Edition (1989) based on the *Complete Works* edited by Stanley Wells and Gary Taylor (1986), the transcriptions of the texts of individual Folio plays in the Oxford Text Archive (OTA) maintained, images of the pages of the First Folio, 1500 works of *Hamlet* art and illustration, several digitized Hamlet films.

MIT's Global Shakespeare Project directed by Peter S. Donaldson has many segments categorised under heads like The Global Shakespeare Video and Performance Archive, Shakespeare in Asia, Educational Videos, Media Annotation Tools, Shakespeare Electronic Archive, Affiliated Projects and Hamlet on the Ramparts. It is a growing archive providing access to international performances that change how Shakespeare's plays are understood. Its online access to Shakespeare performances, essays, and metadata from across the world can help to create newer forms of cultural exchange through the participatory multi-centric networked model. It also leads to global, regional, and national portals to Shakespeare productions and celebrates Shakespeare as a global author. By looking at the diverse ways of his world-wide reception and production, it exploits the possibilities of the digital age and catalogues nearly 300 productions and 75 video clips besides online videos of over 30 full productions.

The richness of the creativity in Asian Shakespeare experiments combining traditional and contemporary theatres, strategies for negotiating languages and genres and their new ways to reach diverse audiences. Have enhanced the remarkably long Asian engagement with Shakespeare. MIT's Shakespeare Performance in Asia hosts Shakespeare videos, photos, and texts in an attempt to promote cross-cultural understanding. It lists some 200 productions and video clips from major Asian Shakespeare productions

Universities are now building digital archives and are moving into new forms of digitally assisted pedagogy to study of Shakespeare and his performance using video clip sets across media, culture, and time. It helps students to experience Shakespeare as a text that has rich performance tradition across the world and enable student to engage with cultures and traditions across the world by working with Using the Internet Folio text of hk's plays from the Internet Shakespeare Edition edited by David Bevington and adding links to corresponding video sequences by scenes, universities create new video library course material by linking URLs already available on the net. It then compares key moments in the play by looking at how these are performed in films and theatres in different languages, cultures and times. MIT Global Shakespeare in Performance modules on Hamlet, The Tempest, King Lear, and The Merchant of Venice using videos of Shakespeare from across the globe. MIT has also linked Shakespeare films available on laserdisc with electronic texts has linked 38 complete famous productions of Shakespeare.

Another intervention in a digital computer is worldwide collaboration. For example, the Project Guttenberg texts are considerably improved through voluntary proof-readers. POS tagging, lemmatization, and proofreading of digital text happen in the online presentation of Shakespeare (Galey).

Funded jointly by UK Joint Information Systems Committee and the US National Endowment for the Humanities, The Shakespeare Quartos Archive is a free digital archive of the pre1642 editions of Shakespeare's plays. It gives at least one copy of every edition of Shakespeare's plays printed in quarto before the theatres were closed by the Puritan Parliament in 1642. This site is intended for the detailed study of quartos which are kept in geographically distant places. Already it has all the 32 quartos of

Hamlet contributed by partner institutions like the Bodleian, the British Library, the University of Edinburgh Library, the Folger Shakespeare Library, the Huntingdon Library and the National Library of Scotland.

Although from a mathematical perspective, little has changed, the lower operational cost in terms of money, time and expertise it is likely that such changes in the digital age are likely to increase and link distant resources virtually and making Shakespeare truly international and making his presence across cultures

Analysing a corpus of 100-million words would have required a dedicated computer facility and manpower when I started to research. Today, even your laptop can process such a corpus in a few seconds, and hardly requires training. Paradoxically, this has a flipside. Just because we manage data faster using a computer, we tend to erroneously think that we understand the text better than we do. The visualised presentation of data can also create a sense of imagined command over the subject and other digitally handicapped scholars, who might be better at understanding Shakespeare than in manipulating data and presenting it in more communicable format.

Conclusion

Is digital Shakespeare the end of literature? *StageGRaph* a digital programme to predict the genre of a play by analysing the different variables found that Othello to become a comedy (Ramsay). Two earlier studies which inverted its comic character also had thought *Othello* must have been a comedy (Stewart; Rogers). *StageGRaph* which arrived at a similar conclusion from a different route seems to validate it.

Digital Humanities use digital technology to understand literature which operates in the realm of the undefined and of possibilities than of the defined and empirical—the realm of science. The large Shakespeare corpus helps science to verify hypotheses.