

# Visualization Analysis of Shakespeare Based on Big Data

Ran Congjing<sup>1</sup>, Li Xinlai<sup>1</sup>, and Huang Haiying<sup>1</sup>

Center for Studies of Information Resources, Wuhan University, Wuhan, China  
School of Information Management, Wuhan University, Wuhan, China

**Abstract** - For more than four centuries, Shakespeare's works have brought great influence on the literary creation around the world. The analysis and digging of strength, distribution, and hotspots of Shakespeare research have helped to understand the current situation and features in Shakespeare's research field, which can provide references for the related research. Based on the theory of big data, information metrology and co-word analysis, this paper makes a visual analysis of the Shakespeare research literature in WOS database. According to the index of documents amount, citations amount and average citations amount, the core authors, core institutions, distribution of journals and research hotspots are analyzed, data and analysis are provided for world Shakespeare's related research from the perspective of knowledge maps.

**Keywords:** Visualization Analysis; Shakespeare Research; VOSviewer; Co-word Analysis

## 1 Introduction

Shakespeare is well known in the world, as his opponent, a contemporary drama writer, Ben Johnson says that Shakespeare does not belong to an era but belongs to all eras. There is no intermittent in Shakespeare's research, or praise or criticism, which shows us a colorful Shakespeare as Shakespeare study walked all the way. On the stage, Shakespeare's drama continued to be interpreted in different forms and languages throughout different countries and regions. In the second decade of the 21st century, the study of Shakespeare continued to flourish and shows diversity of innovation in the academic circles. Shakespeare's research has gone through 400 years, experienced the change eras of various literary trends like Neo-Classicalism, Enlightenment, Romanticism, Realism, Postmodernism and so on. So Shakespeare's research has a wide range of fields, including history, aesthetics, sociology, ethics, philosophy, anthropology, psychology, ecology, film, television art and so on.

But for now, the scholar's research on Shakespeare is mainly based on qualitative research and comparative analysis, and there are very few articles through the pattern of visualization. In order to further study the dynamics of Shakespeare research in the world, this paper adopts the mixed method of quantitative and qualitative, based on the theory and technology of big data analysis, taking the Shakespeare research literature of WOS database from 1900 to 2016. Through the visualization analysis of the amount of literature, the distribution of documents, the citations, and the

key words of the literature, this paper explores the general situation and research hotspots of Shakespeare research abroad, and presents a new research method and references for scholars.

## 2 Methods

### 2.1 Data Sources

The data in this paper is taken from the Web of Science database. In the WOS default core collections, word "Shakespeare" and Shakespeare works' name are used to build the advanced search strategy (appendix 1), with the time span from 1900 to 2016. The retrieval was made on January 3, 2017 (data update date is January 2, 2017) and a total of 35927 data were obtained. Irrelevant data were removed manually under the eliminating standard, a total of 28189 data were effective, including full records and references cited.

### 2.2 Research Methods

Co-word analysis method is one of the important methods of metrological analysis and content analysis method, which mainly reflects the number (or intensity) of word pairs in statistical documents, the degree of close relationship between the words and the relationship between the subject and the theme. Based on the close relationship between word pairs in the keywords network, this paper analyzes the research hotspots and research trend of Shakespeare's research field by using co-word analysis method.

Data visualization can translate data into static or dynamic images or graphics using the basic principles of computer science and graphics, and allows users to interact and control data extraction and display. So hidden knowledge can be dug out and new rules can be discovered. This article mainly uses VOSviewer's visualization technology to draw the knowledge maps. VOSviewer is a software tool for constructing and visualizing bibliometric networks. These networks may for instance include journals, researchers, or individual publications, and they can be constructed based on co-citation, bibliographic coupling, or co-authorship relations. VOSviewer also offers text mining functionality that can be used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature.

### 3 Analysis and Results

#### 3.1 Analysis of Authors' Influence

Influential authors can influence the development trend and is the epitome of the scientific research activities of a certain subject. Through the analysis of the author, it is more easily to grasp the breadth and depth of the development of a subject and have positive significance to the management, organization and coordination of the scientific research activities. The author's academic influence can be measured by the amount of documents and the cited frequency. This paper mainly analyzes the author's influence from the perspective of the documents amount and the average citations.

The paper sets up the relevant parameters in VOSviewer software to analyze author cooperation of the Shakespeare study literature from 1900 to 2016, with a total of 11291 authors, of which 430 authors have more than 5 documents and more than 5 cited times. Visualization map (software automatically filter out "Anonymous") is shown in Fig. 1, the deeper the red, the more the documents, which issued that top five authors are SMITH, PJ, WELLS, S, BERRY, R, WARREN, R, DUNCAN-JONES, K, these researchers are more active and have more prominent contribution in the Shakespeare study.

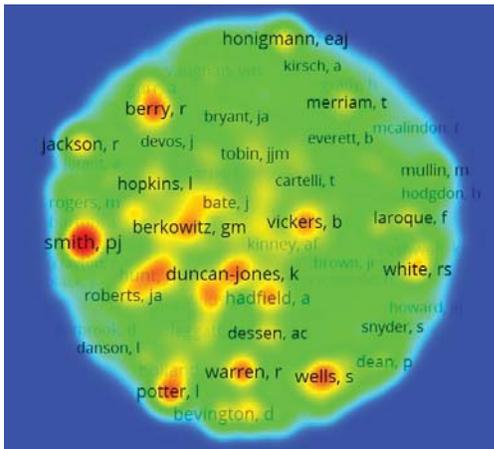


Fig. 1 Literature Volume of Authors in Shakespeare Research 1900 - 2016

The amount of documents is an important measure of authors' academic level and scientific research ability, and we can determine the core scholars of Shakespeare study according to their documentation volume. Those authors who have attained deeper research in the study of Shakespeare are academic leaders. This paper determine the core scholars based on the formula written by Price<sup>2</sup> in Yale University,

$$N = 0.749(\eta_{MAX})^{1/2}$$

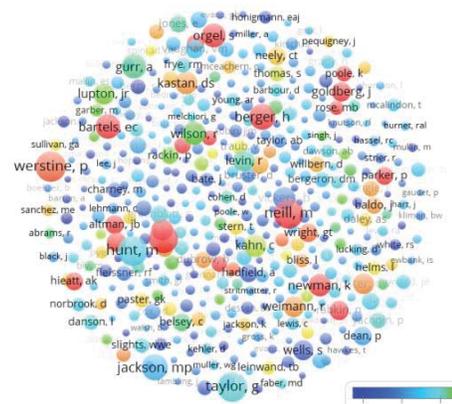
N is the number of papers,  $\eta_{MAX}$  is the number of papers of the author who have the highest yield in the statistics period. Only the author whose published papers amount is more than N can be called the core author. Take the value of  $\eta$  into the formula, calculate  $N = 11$  (articles), so authors who published more than 11 (including 11) papers belong to the kernel

scholars of Shakespeare study. According to statistics, the core scholars of Shakespeare study have 413 authors, following statistics for the top 20 of the author, see Table 1.

Table 1 Documents and Citations of Authors 1900 - 2016 (top 20)

NUMBER	AUTHOR	DOCUMENTS	CITATIONS
1	Smith, PJ	225	17
2	Duncan-Jones, K	145	56
3	WELLS, S	120	51
4	BERRY, R	119	29
5	WARREN, R	109	12
6	PEARCE, GM	95	0
7	MAGUIN, JM	94	4
8	Duncan-Jones, K	92	26
9	MEHL, D	87	2
10	Vickers, B	82	64
11	Vickers, B	82	64
12	JACKSON, R	81	17
13	BERKOWITZ, GM	78	9
14	Hadfield, A	77	41
15	WHITE, RS	77	16
16	HONIGMANN, EAJ	71	21
17	POTTER, L	70	9
18	WILDS, L	70	8
19	JACKSON, MP	68	90
20	HUNT, M	68	119

The quality of the research papers of a scholar can be evaluated by average citations. VOSviewer software can make visualization maps, Fig. 2, of the average citations of above 413 authors by setting citations as weights and average citations as scores. The size of the node represents the amount of citations, and the deeper the red, the more the average citations, and the connection between the nodes represents the cooperation between the authors. Through the relevant parameters in the graph, it can be found that the author with the highest citations is STALLYBRASS, P, the average number of cited is 14.50, followed by BERGER, H, BARTELS, EC, RABKIN, N, NEWMAN, K, the average number of cited are 10.11, 10, 9.6, 8.3, 8 respectively. These authors with low documents amount but high quality are comparatively authoritative in the study of Shakespeare and make great contribution to Shakespeare research as well. In Fig. 2, there is no obvious cooperative group, indicating that the study of Shakespeare is mainly based on individuals. There is no core group between scholars currently.

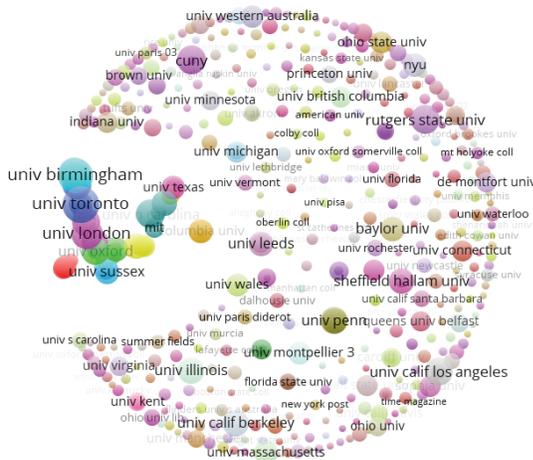


**Fig. 2** Average Citations of Authors in Shakespeare Study 1900 - 2016

12	UNIV ILLINOIS	60	33	0
13	UNIV CAMBRIDGE	59	52	9
14	SHEFFIELD HALLAM UNIV	56	16	1
15	UNIV GEORGIA	55	39	65
16	UNIV CHICAGO	55	42	7
17	UNIV CAPE TOWN	53	25	5
18	HARVARD UNIV	51	108	5
19	UNIV SUSSEX	51	30	2
20	UNIV LEEDS	51	6	1
21	UNIV MARYLAND	51	35	1

### 3.2 Analysis of Organizations' Cooperation

VOSviewer can generate network maps of organizations, with a total of 2160 institutions, of which 475 institutions have more than 5 documents. Visualization analysis is made for institutional cooperation of these organizations, as shown in Fig. 3, in which each node on behalf of an organization, and node's size indicates the number of documents issued by the organization. Organization's influence can also be measured by documents amount and cited amount. This paper mainly reveals the influential organizations in the field of Shakespeare's research from two aspects: published documents and average citations. The top 20 organizations' statistics, table 2, shows that the University of Toronto issued a maximum of 120 papers in the Shakespeare study, followed by the University of Birmingham, a total of 105 documents. In this table, the university is the main force in the field of Shakespeare's research, indicating that academic organizations are the core force of the study of Shakespeare.

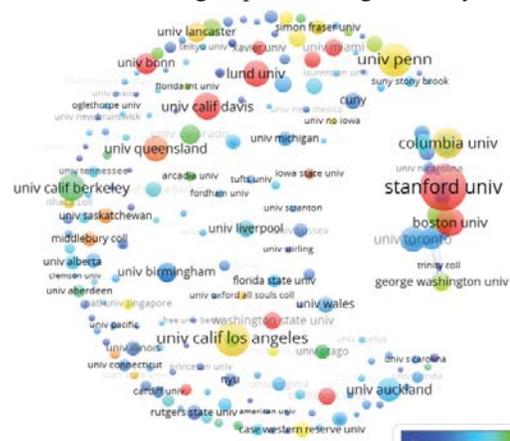


**Fig. 3** Network Map of Organizations in Shakespeare Research 1900 - 2016

**Table 2** Documents and Citations of Organizations in Shakespeare Research 1900 - 2016 (documents > 50)

NUMBER	ORGANAZATION	DOCUMENTS	CITATIONS	TOTAL LINK STRENGTH
1	UNIV TORONTO	120	100	3
2	UNIV BIRMINGHAM	105	62	7
3	UNIV LONDON	85	80	8
4	UNIV N CAROLINA	74	23	64
5	UNIV WARWICK	73	14	3
6	CUNY	73	43	2
7	KINGS COLL LONDON	68	28	7
8	UNIV CALIF LOS ANGELES	68	182	2
9	UNIV PENN	66	170	1
10	UNIV OXFORD	65	54	11
11	UNIV CALIF BERKELY	61	114	0

The visualization analysis, Fig. 4, carried out 258 organizations' average citations with the documents more than 5 and the citations more than 5 by setting the citations as weight. Size of the nodes represents the amount of citations, and the deeper the red, the greater the average citations, and the connection between the nodes on behalf of the cooperation between the organizations. We can find that the University of Lund has the highest citations with the average cited of 20.8 through the following map. This organization has only 5 documents, but the average number of citations is high, indicating that the quality of the published literature is very high of this school. Followed by the University of Bonn, Stanford University, with the average citations of 11.67, 10.38 respectively. Some institutions have papers of high quality, even though with low documents amount can be found by the index of average citations. As it can be seen from the map below, there is little co-operation between most institutions. But there is a cluster including 47 universities, such as Stanford University, Boston University, Columbia University, Toronto University, George Washington University, University of Georgia, MIT, Harvard University, University of Notre Dame, University of Oxford, University of California, San Diego and so on, which shows that university are the mainly organizations to make cooperation in Shakespeare's research, and a research group is forming currently.



**Fig. 4** Average Citations of Organizations in Shakespeare Research 1900 - 2016

### 3.3 Analysis of Journals' Distribution

After the statistical and visual analysis of the source of the documents of Shakespeare research, the number of



The word "Hamlet" is not only the name of Shakespeare's a work, but also the name of the hero in this famous work, so in the map the largest node shows that Hamlet is the keyword which have most co-occurrence times, and 84 other keywords appeared in the co-occurrence, so the study of Hamlet took a major share. These documents study the Hamlet from different angles, not only from the respect of characters, humanism, and madness, some studies also try to analyze it from the perspective of religion, Postmodernism and Postmodern. Hamlet is a tragic work written by Shakespeare from 1599 to 1602. Keyword "Ophelia" is one of only two female roles in Hamlet, and is the lover of Hamlet, so the times of co-occurrence with Hamlet is in great quantity. As one of Shakespeare's four tragedies, both from the story itself or from the Hamlet's character, it would be related with words like "tragedy" "sadness" "mourning" and other keywords.

The genre of King Lear is drama, which comes from an ancient legend of the United Kingdom. The story itself occurs around the 8<sup>th</sup> century, which was adapted into a lot of drama in British after that. There is an earlier anonymous work in the existing versions except Shakespeare's drama, so it is generally believed that Shakespeare's King Lear is the adaptation of that play. The place of that story is the United Kingdom and France, and the time is around the 8<sup>th</sup> century, which with the background of Renaissance when ideas of humanism and new historicism continue to develop and spread. So the research and interpretation for King Lear have aspects from humanism, new historicism, and therefore associated with the keywords like "humanism", "new historicism" etc. Also as one of the four tragedies of Shakespeare, it is common to study King Lear from the tragic point in many papers, such as the analysis of the four tragedies, the formation of the Lear King's tragedy, the tragedy color from the perspective of literature, etc. so the keyword "tragedy" generated high contributions to the frequency of co-occurrence.

### 3.4.2 Studying Shakespeare Works from the Perspective of Theme

The second Shakespeare study hotspot cluster is the analysis of Shakespeare's famous works from the perspective of the theme, like the famous characters, the thematic diversity and so on. All Shakespeare's tragedy, comedy and historical drama reflect the diversity of its theme. And in Shakespeare's works, he shapes a lot of typical images, such as female image, king image, statue, etc. Therefore, lots of literature excavate and analyze the theme of Shakespeare works, such as "time" theme and "love" theme in the Sonnet, "death" theme and "revenge" theme in Four Great Tragedies. This paper chooses the two outstanding works "Othello" and "Venice Merchant" as examples.

Othello is also one of the four tragedies, written about in 1603. The study of this work is mainly from the perspective of the theme of the work, such as character, racial discrimination and colonialism, feminism and so on. "Othello" is a multi-themed work, including the theme of love and jealousy, the

theme of credulity and treachery, the theme of intermarriage, and racial discrimination is one of the reasons for the tragedy of Othello. Shakespeare succeeded in exposing some features of his time, like the theme of racism, through the description of the whole process of Othello's failure, a person from different nation<sup>4</sup>. There is a conflict between the African-American culture presented by Othello and the European mainstream culture presented by Desdemona, and it is helpful for today's multi-cultural exchange and integration through the study of this theme<sup>5</sup>. In addition, there are some papers make the visualization analysis of the version of Othello<sup>6</sup>, and the comparative research between Othello and "Hamlet", "King Lear", "Macbeth"<sup>7</sup>, so it is related with the keywords "translation", "Hamlet" "Macbeth" "King Lear" and have more co-occurrence times with these words of co-occurrence.

The Merchant of Venice is a great irony comedy, written about 1596-1597 years. The theme is love and friendship, but also reflects the contradictions between early commercial bourgeoisie and usurers in capitalist society, which expresses humanism through the problems about money, law and religion in the bourgeois society. An important literary achievement of this play is to shape the typical image of Shylock<sup>8</sup>, the ruthless usurer, and many papers analyze this work from the points of characters and racism, so keyword "The Merchant of Venice" related with keywords "Shylock", "Race" more often. In addition, other common keywords, such as "Macbeth", "Midsummer Night's Dream" are also compared with high frequency.

### 3.4.3 Studying Shakespeare Works from the Perspective of Textual Criticism

The content of this cluster is mainly about author's identity of Shakespeare's work. As drama script belongs to the theater in Shakespearean Age, there is no signature on the Shakespeare's early drama scripts, and people have no awareness of copyright, resulting in a lot of works lost and the author can not be distinguished. At the same time, due to creation and adaptation of scripts in Shakespearean era, it is common for the phenomenon of cooperation between authors, which can explain that the keywords with high frequency of co-occurrence including "collaboration". For the above reasons, many authors have a discussion and controversy for some dramas' identity and the situation of cooperation, including Shakespeare's lost works, false works and suspected works, such as "Love's Labour's Won", "Cardenio", "The London Prodigal", "Sir Thomas More", "Arden of Faversham", and so on.

Another content of this cluster is the study of Shakespeare's adaptation works. The study of the other forms like films, plays, dramas, TV shows adapted from Shakespeare's text works, such as Fortier, Mark studied the adaptation of Shakespeare's five plays<sup>9</sup>, Jones, Keith studied the relationship between the text of Shakespeare's Hamlet and films<sup>10</sup>, Mike studied the music adaptation of Shakespeare's sonnets. The study of the adaptation of Shakespeare's works for different people, such as the adaptation for adolescent groups<sup>11</sup>; the limitations of religion, nation, gender and

adaptation of Shakespeare's work for the specific country or region, like Canada, China<sup>12</sup> and so on.

## 4 Conclusions

With the 400th anniversary of the death of Shakespeare, there would be another climax of Shakespeare study. This paper summarizes the research situation and hotspots of the global Shakespeare research by statistical analysis and visualization analysis of the research literature, mainly including the following findings:

Shakespeare's research is mainly based on individual authors, and there is no core group of authors currently. Academic institutions, mostly universities, are the core of the study of Shakespeare. Journals are widely distributed but the research field is relatively concentrated on the field of literature. There are three main hotspots in the study of Shakespeare's research literature: research Shakespeare's famous works from the point of history like historical drama genre or historical background etc., research Shakespeare's famous works from the perspective of the theme like characters' image, characters' personality or theme and so on; and research Shakespeare's works from the perspective of textual criticism including the study of authorship, Shakespeare's doubt works, Shakespeare's cooperating works, Shakespeare's adapted works and so on.

## Acknowledgment

This work is supported by the project Intellectual Property's Institution Change, Risk Analysis and Countermeasure Research of Information Consumption in Cloud Environment of Humanity and Social Science Research Project of Ministry of Education of China. (Grant No. 14YJA870008 )

Ran Congjing, Li Xinlai and Huang Haiying contributed equally to this paper and should be considered co-first authors.

Li Xinlai is the corresponding author of this paper.

## References

- [1] VOSviewer.Home[EB/OL]. <http://www.vosviewer.com/>,2017-01-13.
- [2] Zhu Y. The Identification and Evaluation of the Kernel Authors of New Technology of Library and Information Service[J]. *New Technology of Library & Information Service*, 2004.
- [3] Bradford S C. Sources of Information on Specific Subjects[J]. *Journal of Information Science*, 10(4):173-180, 1985.
- [4] Smith, I. We are othello: speaking of race in early modern studies. *Shakespeare Quarterly*, 67(1), 2016.
- [5] Nachit M. Shakespeare's Othello and the Challenges of Multiculturalism[J]. *Social Science Electronic Publishing*, 2016.
- [6] Zhao, G., Cheesman, T., Laramée, R. S., Flanagan, K., & Thiel, S. Shakervis: visual analysis of segment variation of German translations of shakespeare's othello. *Information Visualization*, 14(4), 2013.
- [7] Davis, A. Shakespearean tragedy as chivalric romance: rethinking 'Macbeth', 'Hamlet', 'Othello', 'King Lear', by Michael I. Hays. *Modern Language Review* (1), 226-227, 2006.

- [8] Bailey, A. Shylock and the slaves: owing and owning in the merchant of venice. *Shakespeare Quarterly*, 62(1), 1-24, 2011.
- [9] Johanson, K.. Shakespeare adaptations from the early eighteenth century : five plays. Fairleigh Dickinson University Press, 2013.
- [10] Nelson L M. Screen Adaptations: Shakespeare's Hamlet: The Relationship between Text and Film by Samuel Crowl (review)[J]. *Journal of Dramatic Theory and Criticism*, 31(1): 174-176,2016.
- [11] Mike Ingham. "the true concord of well-tuned sounds": musical adaptations of Shakespeare's sonnets. *Shakespeare*, 9(2), 220-240, 2013.
- [12] Drouin J. Shakespeare in Quebec: Nation, Gender, and Adaptation[M]. University of Toronto Press, 2014.

## Appendixes

1 Retrieve formula: TS=(Shakespear\* OR Shakesperian OR "William Shakespeare" OR "All's Well That Ends Well" OR "As You Like It" OR "The Comedy of Errors" OR "Love's Labour's Lost" OR "Measure for Measure" OR "The Merchant of Venice" OR "The Merry Wives of Windsor" OR "A Midsummer Night's Dream" OR "Much Ado About Nothing" OR "Pericles, Prince of Tyre" OR "The Taming of the Shrew" OR "The Tempest" OR "Twelfth Night" OR "What You Will" OR "The Two Gentlemen of Verona" OR "The Two Noble Kinsmen" OR "The Winter's Tale" OR "Cymbeline" OR "King John" OR "Edward III" OR "Richard II" OR "Henry IV" OR "Henry V" OR "Henry VI" OR "Richard III" OR "Henry VIII" OR "Romeo and Juliet" OR "Coriolanus" OR "Titus Andronicus" OR "Timon of Athens" OR "Julius Caesar" OR "Macbeth" OR "Hamlet" OR "Troilus and Cressida" OR "King Lear" OR "Othello" OR "Antony and Cleopatra" OR "The Sonnets" OR "Venus and Adonis" OR "The Rape of Lucrece" OR "The Passionate Pilgrim" OR "The Phoenix and the Turtle" OR "A Lover's Complaint" OR "A Funeral Elegy" OR "Sonnets to sundry notes of music" OR "Love's Labour's Won" OR "Cardenio" OR "Sir Thomas More" OR "Arden of Faversham")