A Study of the Application of Word Cloud Visualization in College English Teaching

Yichang Huang, Ye Wang, and Feng Ye

***Abstract*—Nowadays, the cultivation of students' reading ability has become one of the essential aspects in College English Teaching, in which the methodological guidance to help students grasp effectively the key words of articles emerges particularly as the significant point in classroom interactivity. As the visualization mechanism analyzing and researching the English text quickly, Word Cloud has been put forth as English teachers' new tool in reading training. This thesis tends to present a dynamically interactive system of Word Cloud Visualization based on the technique of D3 and JQuery in the field of Computer Science, and attempts to elaborate its advantages, such as the comfortable interface and the prominent key points with several examples of recent US presidential inaugural speeches as well as the test analysis of the CET 4 .**

***Index Terms*—College English teaching, D3, visualization, word cloud.**

I. Introduction

At present, College English is an important compulsory course at the undergraduate level in China. It's significant to help students grasp the key words of English articles proficiently and efficiently, furthermore, it aims to improve students' ability of reading comprehension tremendously in the classroom interactivity of College English Teaching. As the era of big data [1] is coming, visualization [2] techniques have also achieved remarkable contributions in expression and presentation of texts, networks or pictures, space-time and multidimensional data. Among all these novel prospects, the representative techniques grounded with English word frequency is Word Clouds [3], which is theoretically applicable to text visualization.

A word cloud, also known as a text Cloud or tag Cloud, sorts key words of English articles according to the word frequency or other rules, arranging them based on fixed laws, and visualizing them with graphic attributions such as size, color and font. In general, once having higher frequency or showing more importance among a certain group of words, the word cloud would highlight the more prominent words in a direct and visible way. In the Internet application with prevailing graphical and superficial reading, the value of word clouds lies in screening, analyzing, interpreting the seemingly disordered textual data and providing the focus of text, as well as revealing important concepts and penetrating into the nature of information previously inaccessible by the new pattern. The word cloud can make rapid and efficient analysis and research on the text, even towards a variety of texts, to help readers understand the connotation and theme from different perspectives. Since most of the core content of College English Teaching are reading materials organized by text data, it is clear that the mechanism of Word Clouds Visualization could become a good auxiliary tool in the current practical classroom interactivity.

College English Teaching research based on the word cloud [4]-[8] has drawn wide attention and achieved some success. However, there is still some space for improvement on the grounds of word cloud visualization technique, which is mainly reflected in the following aspects. For instance, word cloud visualization tools represented by Wordle [9] and Tagxedo [10] can rapidly display text analysis or word frequencies on the website of various styles, and support font choice and user-defined colors, the flexibility, immediacy and interactivity of network software in the third-party remain lackluster. As usual, in the certain English teaching activities, students from different classes and grades harvest widely different effects when learning the same article or words. In this case, word clouds need to adjust dynamically according to students' mastery of knowledge or their points of interest in language acquisition. Therefore, this thesis attempts to put forth a dynamic and interactive mechanism of Word Cloud Visualization, based on the technique of D3 [11] and JQuery [12], in order to enrich English teaching methods and improve the learning effectiveness.

II. Advantages of Word Cloud Visualization in College English Teaching

Within the scope of reading practice in College English Teaching, the articles are usually organized by words based on textual description and grammar. Such linear sequences have primarily tended to solidify the cognitive process and brought about certain negative impacts on English learning efficiency. Nevertheless, the presentation of word cloud visualization cannot be limited by the current sequences, whose graphical information organization model is efficient for students to understand the meaning of the whole text within a comparatively short period of time. Word clouds are designed according to the basic characteristics of human visual memory, that is, memorizing images is always superior to memorizing words. Henceforth, when dealing with complicated English articles, word clouds could completely mobilize the English words stored in the human memory bank, and stimulate students to transform the articles into visual and memorable texts, thereby instantly improving their efficiency of English reading.

Manuscript received April 12, 2018; revised August 23, 2018.

Yichang Huang and Ye Wang are with College of Foreign Languages, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China (e-mail: clarisse2015@163.com, 2571258244@qq. com).

Feng Ye is with College of Computer and Information, Hohai University, Nanjing, Jiangsu, China (e-mail: yefeng1022@hhu.edu.cn).

doi: 10.18178/ijiet.2019.9.2.1185

In traditional teaching of English reading, reading materials are usually only be remembered according to the arrangement of lexical linear sequences, which makes it more difficult for students to understand the content of a certain article within a short time. Word clouds endeavor to change the innate feature of word vision to strengthen the visual memory with highlight of key words. Generally, substantive words are the key words of an article. Once grasping the elemental substantive words, students could easily recall the content of a certain article, which naturally reduces the obstacles caused by linear sequences originally existing in the articles. Meanwhile, words in the word cloud are high-frequency vocabulary in the reading materials, which are embedded with a sense of familiarity in students' memory to some extent. In the light of the font size, color and so on, students can understand the key words quickly and further have a good command of the meaning of the article.

With the visible advantage in reading practice, word clouds could be further applied into the wider range of College English Teaching. Nowadays, it remains to be difficult for students to distinguish the words with similar forms, not only in daily English learning, but also in some important national English tests such as CET4 (College English Test Band 4) and CET6 (College English Test Band 6). If teachers could generate the corresponding word clouds according to the proportion of wrong words in students' exercises, tackling with the main reasons of making mistakes, and then dynamically adjusting word clouds based on the teaching feedback, this approach would be obviously conducive to enhance students' learning efficiency of English words, and moreover, avoiding making similar mistakes and improving their passing ratio in the national tests.

III. Implement of Word Cloud Visualization

Due to the implement of Word Cloud Visualization mechanism, firstly, users need to remove the non-alphabetical symbols and spaces in the English articles selected by using the regular expression "[^ a-zA-Z]." In the next ground, based on the data structure of Red-Black Tree [13], the word-frequency statistics of words in these articles are respectively calculated [14], is which the algorithm is shown in Fig. 1 below. Consequently, the well-handled articles and word-frequency statistics have been stored in the back-end database. For some unimportant or infrequent words, such as "we," "of," "the," etc., users can set appropriate filtration mechanism to highlight the keywords in a more transparent manuscript.

Secondly, taking advantage of D3-Clouds, the realization of open-sourced word clouds' layout based on D3.js library, vividly exhibits the function of Word Cloud Visualization. The application of Slider Control of J Query-UI could arrange word clouds of different English articles in a chronological order, for instance, in the word cloud (Fig. 2) from President George W. Bush's inaugural speech in 2005, the system's user interface is shown clearly. Furthermore, the front-end word cloud could display interacts with the back-end database by using JSON data format.

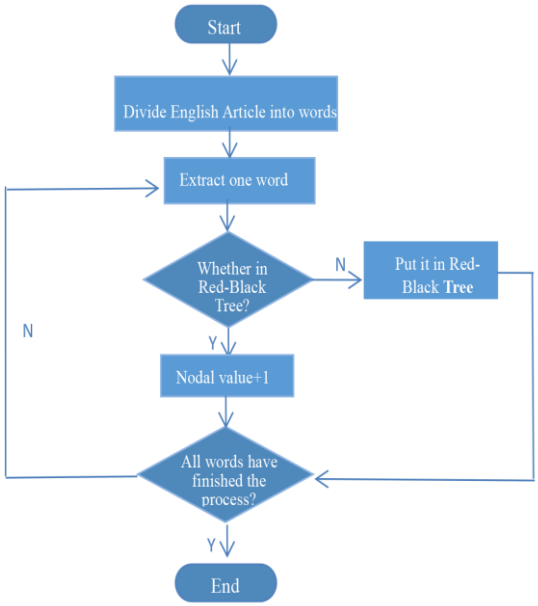


Fig. 1. Flow chart of word-frequency statistics.

Fig. 2. User interface of word cloud visualization software.

IV. Cases and Effects

In order to elaborate the function of word cloud, particularly, the effectiveness of D3 and J Query in reading practice, the inaugural speeches of recent two U.S. Presidents are researched as case studies, namely, the inaugural speech of President George. W. Bush in 2001 and that of President Barack Obama in 2013. First of all, the speech of Bush has been made automatically on the basis of word frequency, as is shown in Fig. 3, in which the mechanism of Word Cloud Visualization is demonstrated and analyzed clearly. According to the preliminary analysis, word clouds are directly disturbed by word frequency, which can not distinguish the part of speech, in this case, such pronouns as "we" and "our" are highlighted because of the frequent appearance. however, these words do not include practical or pragmatic meaning in understanding the text. Therefore, the technique of D3 and J Query is introduced into the process of making up a new word cloud, as is shown in Fig. 4, some important words, such as "freedom," "America," "work" and so on could be gradually highlighted with the adjustment of the filter. Undoubtedly, the new word cloud becomes more clear and prominent in presenting the core meaning of Bush's

speech with which students would acquire self-esteem and sense of achievability. In other words, the D3 and J Query technique is quite essential to distinguish automatically the substantive words and the functional words, which could display a more transparent word cloud to be easily understood.

Word cloud can not only visibly show the theme of the speech in a short time, it further accelerate the comparison process of these two speeches. Through adjusting the space of time axis, emphasis in different U.S. presidential speeches could be compared and contrasted clearly. Fig. 5 is a filtered word cloud based on President Barack Obama's speech. Obviously, contrasted with Fig. 4, different word clouds make readers understand differences on the themes or the emphasis of speech easily, as long as users repeatedly adjust words which need to be filtered and then generate the required images dynamically and instantly. In addition, once the word clouds are made, the essential theme of the speech become transparent, which exactly directs to the similarity and difference of central innovation of each President. For example, both in George W. Bush and Barack Obama's speeches, the frequently mentioned words are "freedom", “America/American”, “country”, “us”, etc., which illustrate the traditional feature of inaugural speech as a custom. It seems that both of the American Selected Presidents prefer to emphasize the spirit of pursuing freedom and the connotation of country, to call on people to follow their upcoming reforming policies as a united group. Meanwhile, according to these two word clouds (Fig. 4 and Fig. 5), the difference of certain frequently-appeared words uncovers the diversity of the innovation orientation between these two presidents and discovers the change over the decade. Obviously, President Bush has put more focus on practical solutions related to “school” and “society”, which is more inseparable to people's daily life, while President Obama tends to encourage people with the metaphor of “journey” and describe the future prospect with some verbs such as “believe”, there is no denying that the former's speech is more pragmatic and functional in terms of policy making. Through the comparison mentioned above, students are not only familiar with the contents of these speeches, they are more clear about the stylistic feature of them, in other words, the word clouds help students to comprehend the central meaning in reading practice and strengthen their capacity of literary appreciation which belongs to the advantageous step of English teaching.

Except for the application into the practical daily reading training, the word cloud could also be used in English test to help students immediately understand the meaning of passages and save time in real test. Such method used to be practiced in the writer's English class and produces great effect. Concretely, the teacher uses the first passage of the reading comprehension in the test of CET 4 (2017.06) as follows: “As if you needed another reason to hate the gym, it now turns out that exercise can exhaust not only your muscles, but also your eyes. Fear not, however, for coffee can stimulate them again. During 26 exercise, our muscles tire as they run out of fuel and build up waste products. Muscle performance can also be affected by a 27 called “central fatigue”, in which an imbalance in the body’s chemical messengers prevents the central nervous system from directing muscle movements 28. It was not known, however, whether central fatigue might also affect motor systems not directly 29 in the exercise itself, such as those that move the eyes”. [1] In this paragraph, students are asked to fulfill the blanks from 26 to 29 with the given multiple words, respectively, “cautiously”, “commit”, “control”, “cycling”, “effectively”, “increased”, “involved", “limited”, “phenomenon”, “preventing”, “sensitive”, “slowing”, “solution”, “sufficient”, “vigorous”. Traditionally, this reading comprehension is always the difficult part in CET 4, because students would grasp the main idea at the first reading step so as to judge which word is more suitable to the blanks at the second close reading. To testify the function of word cloud in reading ability training, 20 freshmen with similar basis of English skills in the same class are selected as the target to be tested. First of all, the students are distributed into two groups, one is following the traditional way of practicing reading comprehension, the other is guided to consult a word cloud containing the key words of the paragraph before reading. Later on, these two groups are required to scan the whole article within 10 minutes, and then fulfill these blanks. When checking the complement of the task, there is an obvious result, that is, the first group who directly get in touch with the article get a lower ratio of correctness while the second group almost get 100% correctness even if they cost more time at the very first step of being familiar with the word cloud. In the filtered word cloud, such substantive words such as “exercise”, “performance”, “fatigue” exhibit as the highlighted words to disclose the central idea of the paragraph, meanwhile, students could sense the stylistic feature according to the word frequency, which enables them to grasp the coherence of the paragraph and quickly match the blank with the multiple choices in the answers.

As is shown in the examples above, the application of word cloud visualization mechanism does make the teaching of English reading more vivid and interesting, improving students’ ability of English reading comprehension. Furthermore, it could reduce the time of first reading with a fast understanding of the passage. Due to the visible advantages, this teaching method could not only used in the daily reading exercise or model test of CET 4 and CET 6, it further expands in the retrospect of proofreading. Given the analysis of words in examination papers, teachers could generate word clouds based on words’ error-prone degree and their difficulties at each section of the test, which could remind students of avoiding the similar mistake in the following tests, giving them a deeper impression, strengthening their visual memory and bringing higher teaching efficiency. In this sense, this tremendous enhancement of reading capacity via application of word cloud overpasses the traditional and boring way of classroom interactivity of blackboard writing and oral introduction. The visible manuscript of word cloud has tremendously invoked students' interest in English class and acquired satisfying teaching effect in the concrete class interactivities, which is

[1]The original passage of the first reading comprehension in the test of CET 4 (2017.06) is omitted here because of the space limit of the paper. See: http://www.for68.com/siliuji/shiti/sx1709113033.shtml. Accessed on Jan. 2018.

expected to have a wider range of research field and development in future.

Fig. 3. The word cloud of George W. Bush's inaugural speech in 2001.

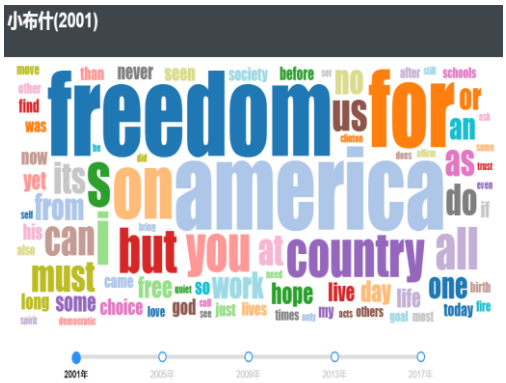


Fig. 4. The word cloud of George W. Bush's inaugural speech in 2001 after adjustment of the filter.

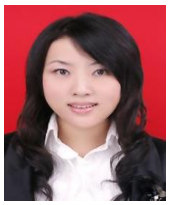
Fig. 5. The word cloud of Barack H. Obama's inaugural speech in 2013 after adjustment of the filter.

V. Conclusion

At contemporary society, Education big data has developed vigorously with a brand new involvement into the traditional College English Teaching, while the research and application of visualization mechanism remains to be a important trend, in which the significance of text visualization lies in directly presenting semantic features (such as word frequency and its importance, logical structure, thematic clustering, dynamic evolution rules, etc.). Combined with big data and visualization technique sufficiently, word clouds inject new vitality to traditional English teaching methods in the current practice of English basic skills training. The dynamically interactive mechanism of Word Cloud Visualization has also achieved good application effects. According to the evaluation of the theoretical expansion and the effect in the practical application, this mechanism is expected to have two research fields in future: (1) to intensively research how to apply the mechanism better in many cases so as to obtain abundant teaching achievements in College English Teaching; (2) to make association analysis of students' learning state and effectiveness, which is further integrated with trajectories generated from different word clouds.

References

[1] G. J. Li, "The scientific value of big data research," *Communications of the CCF*, vol. 8, no. 9, pp. 8-15, Sep. 2012.
[2] R. Lei *et al*., "Visual analytics towards big data," *Journal of Software*, vol. 25, no. 9, pp. 1909−1936, Sep. 2014.
[3] J. Y. Tang and M. S. Sun, "Word cloud in new media: A visualized form with concise content," *Science & Technology For China's Mass Media*, no. 11, pp. 18-19, Nov. 2013.
[4] L. J. Cong, "The application study of word cloud in English vocabulary teaching," *Journal of Nantong Navigation Technical University*, vol. 15, no. 1, pp. 121-124, Mar. 2016.
[5] L. Wang, "The application study of word cloud in English reading teaching," *Journal of Reading and Writing*, vol. 11, no. 6, pp. 51-52, June 2014.
[6] X. Y. Huang, "Application of word cloud in audiovisual multimedia technology of foreign languages," *E-Education Research*, no. 12, pp. 98-105, Dec. 2011.
[7] H. Du and Y. C. Sun, "Application of word cloud in English reading teaching," *China Educational Technology*, no. 8, pp. 115-119, Aug. 2012.
[8] B. Melissa, S. Pennestri, and M. Selvandin, "Using WORDLES to teach foreign language writing," *Language Learning & Technology*, vol. 15, no. 2, pp. 12-22, June 2011.
[9] F. Jonathan. Wordle. [Online]. Available: http://www.Wordle.net
[10] Tagxedo. [Online]. Available: http://www.tagxedo.com
[11] D. Mike, *Getting Started with D3*, Sebastopol: O'Reilly Media, 2012.
[12] S. Eric, *jQuery UI*, Sebastopol: O'Reilly Media, 2012.
[13] Q. Z. Chen, "An efficient binary search tree — Red/black tree," *Journal of East China Normal University (Natural Science)*, no. 3, pp. 39-42, Sep, 2000.
[14] Y. Luo *et al*., "Text keyword extraction method based on word frequency statistics," *Journal of Computer Applications*, vol. 36, no. 3, pp. 718-725, Mar. 2016.



Yichang Huang is currently a lecturer at Nanjing University of Aeronautics and Astronautics, China. She received her PhD degree in Anglo-American literature from Nanjing University in 2014. Her research interests include contemporary British and American fiction, narratology and English teaching.

Ye Wang is currently a MA student in English language and literature at Nanjing University of Aeronautics and Astronautics. She received her B.A. degree in English from Zhejiang Ocean University in 2016. Her current research interests include British and American literature.

Feng Ye is currently a lecturer at Hohai University, China. He received his PhD degree in computer application technology from Nanjing University of Aeronautics and Astronautics in 2016. He is a member of the China Computer Federation. His research interests include dependable computing, cloud computing and big data.