



Research on the Extraction and Application of Vinegar Culture Elements in Zhenjiang Under the Background of Big Data

Gan Gao and Mingzhu Li^(✉)

Department of Industrial Design, Art School of Jiangsu University, Zhenjiang, Jiangsu, China
50884420@qq.com

Abstract. As one of the Four Famous Vinegars of China, Zhenjiang vinegar has a long cultural history and great cultural significance, which is a profound refinement in the lifestyle and cultural spirit of Zhenjiang people. In recent years, with the explosive growth of data resources, the channels of access to information are expanded, and the contents are even more complex. It takes a lot of time and energy to get the content of Zhenjiang vinegar culture quickly and accurately, extract high-quality information resources, and then use them for innovation and culture restoration. In this study, big data technology was used along with traditional research methods to obtain materials for Zhenjiang vinegar culture and identify important cultural elements. It explores the feasibility and cultural innovation approach of applying data processing technology to the design process of a mobile H5 picture book under the background of big data.

Keywords: big data · web crawler · card sorting · Zhenjiang vinegar culture · H5 picture book

1 Introduction

Vinegar, as the most common seasoning in daily cooking, has been around for three thousand years. There are countless records about its history, crafts, diet, myths, and medical uses. Zhenjiang fragrant vinegar is one of the Four Famous Vinegars. According to textual research, the usage of rice vinegar was recorded in the *Bencaojing jizhu* by Tao Hongjing of the Liang Dynasty. It shows that Zhenjiang fragrant vinegar should be the same as the rice vinegar mentioned in Tao Hongjing's book. It has a history of more than 1400 years [1], so it is difficult for us to collect its cultural information. However, with the development of information technology and the popularity of the internet, information has entered people's daily life in digital form. At the same time, sophisticated database technology can store an enormous amount of data for sorting and retrieval, making data sharing and extraction simple and inexpensive [2]. However, the ability of individuals to collect and process data and information is so limited that we must take certain technical measures to help people gather and organize information quickly and efficiently.

Under the background of large amount and various types of information data in the era of big data, this study integrates the data of Zhenjiang vinegar culture obtained by

web crawlers, literature research, and field investigation, and carries on the statistics of Chinese word segmentation and word frequency based on these data, classifies and analyzes the data with the method of card sorting, and obtains the relatively objective and accurate content of Zhenjiang vinegar culture for story creation, to help designers to design a mobile H5 picture book.

2 Materials and Methods

Before collecting data, we need to clarify the content of Zhenjiang vinegar culture to prepare for further data screening. This study aims to guide the design of H5 mobile picture books and promote the spread of culture by using big data technology. Therefore, from the perspective of reflecting cultural phenomena, expressing cultural forms, and interpreting cultural connotation, Zhenjiang vinegar culture is defined as a phenomenon that reflects or represents the material, spirit, or system of Zhenjiang civilization through the organic combination with vinegar, taking vinegar as the medium or carrier [3]. After determining the definition of vinegar culture, we began to collect the data of vinegar culture through three channels.

2.1 Data Collection of Zhenjiang Vinegar Culture

- 1) *Data collection based on web crawler:* Web crawler is a type of technology that can build an index, search and categorize information automatically on a web page, extract web content and retrieve information resources on the Internet quickly and efficiently with a search engine [4, 5]. It obtains the content that the user is interested in by formulating a search strategy, then extracts a new URL from the current page and puts it in a new URL sequence according to the specified order until it meets the system requirements, compares and stores crawl data to perform data collection [6]. In the data collection stage based on web crawler, octopus tool is used to collect the information with the keyword “Zhenjiang vinegar culture”. Data collection is mainly divided into three modules: social platform, news website, and special site. Among various social platforms, we choose WeChat and Weibo because both have a large number of users. A total of 279 data articles with more than 490000 words are retrieved from WeChat, and A total of 498 microblog data with more than 60000 words are retrieved from Weibo. The news websites include Jinri Toutiao, CCTV, People’s Daily, Xinhua News Agency, and other sites, all of which are great, high-quality news sites in China. A total of more than 610,000 words are collected. The website of the Hengshun group, Zhenjiang Intangible Cultural Heritage, and Chinese intangible cultural heritage, which are closely related to the research topic, are chosen as special sites, with a total of more than 90,000 words obtained.
- 2) *Data collection based on literature retrieval:* By searching, collecting, identifying, organizing, and analyzing Zhenjiang vinegar literature, we can gain an objective and relevant scientific understanding of Zhenjiang vinegar culture from our predecessors and analyze this data from literary studies. It can help researchers know about current academic views and lay the foundation for their research [7]. Taking “Zhenjiang vinegar” and “Zhenjiang vinegar culture” as the subject, a total of 50 articles highly

related to vinegar culture are collected in the literature retrieval in China CNKI database, involving history, technology, function, literature, and other aspects. Then the contents of the Zhenjiang vinegar culture in these articles are extracted and reorganized, and over 20,000 words are collected.

- 3) *Data collection based on field investigation*: As a field research method, field investigation can go deep into the life of the respondents, collect first-hand research data by means of observation and interview, and achieve the purpose of specific research through the sorting and analysis of data [8]. In this study, we visit the Zhenjiang Vinegar Cultural Museum, observe the exhibits, and conduct an informal interview with commentators. In total, more than 300 photos are taken and more than 60 min of video and audio are recorded. After the investigation, the image, video, and audio data are sorted and summarized, and more than 5,000 words of relevant information are obtained. The combination of three channels can take full advantage of each method, make up for the shortcomings and collect the data related to Zhenjiang vinegar culture as thoroughly and objectively as possible.

2.2 Data Screening

In the process of data collection, although keywords are set to limit the content of index data, some articles that are not related to Zhenjiang vinegar culture and do not conform to the definition of Zhenjiang vinegar culture are still involved, including but not limited to push ads, business reports, personal travel notes, and other articles. Therefore, to make the word segmentation and word frequency statistics of Zhenjiang vinegar culture data more convenient and accurate, these data need to be screened manually, and the data collected by web crawlers, literature research, and field investigation should be integrated into a document.

- 1) *Jieba Chinese word segmentation*: After the document is integrated, it is still a huge data group. To get the specific cultural phenomena and the corresponding frequency data, it is necessary to conduct word segmentation. The word segmentation tool used in this study is Jieba word segmentation based on Python. Users can achieve word segmentation through three modes: accurate mode, full mode, and search engine mode. In the accurate mode, sentences can be segmented with high precision, which is more suitable for text analysis; the full mode can scan all the words in the data that can form words; the search engine mode can accurately segment long words [9]. Considering the integrated data form and the desired data content, this study uses the accurate mode for word segmentation, and on this basis, adds a discontinuous synonym list to filter out some function words irrelevant to the research content. As there are a large number of professional words, rare words, and book titles in the crawled data of Zhenjiang vinegar culture, these proper nouns can be easily segmented into meaningless words by Jieba, so they should be retained by adding user-defined dictionaries. Finally, the original data of Zhenjiang vinegar culture is segmented into high-precision words by introducing the stop word list and user-defined dictionary through the accurate mode of Jieba word segmentation.
- 2) *Low-frequency word and interference word filtering by Mini Tag Cloud*: After the sentence segmentation, there are still many low-frequency words and interference

words to be removed. Although researchers can filter those words by adding them to the stop word list in Jieba, they have to update the stop word list and filter it repeatedly to achieve the best effect. Mini Tag Cloud is a practical and simple online word cloud generator, which supports online word segmentation, word frequency statistics, and word frequency analysis [10]. In the Mini Tag Cloud, the operation of filtering is different. When we import data segmented into words, the generated entries will be listed one by one. Users can uncheck them in an intuitive list of entries, remove these low-frequency words and interference words, avoid repeating the complex operation of filtering, directly retain the optimized entries and generate word frequency statistics. After completing the above steps, we got 63 entries of Zhenjiang vinegar culture based on big data, numbered 01–63. They are 01 Hengshun, 02 fragrance, 03 brewing wine, 04 solid-state layered fermentation, 05 HeiTa invented vinegar, 06 acid, 07 vinegar jar, 08 stories of tasting the vinega, 09 glutinous rice, 10 vinegar workshop, 11 intangible cultural heritage, 12 acetic acid, 13 color, 14 thickness, 15 Hengshun fragrant vinegar brewing technique, 16 drench the vinegar, 17 solid fermentation substrate of vinegar, 18 mellowness, 19 vinegar protection regulations, 20 Zhenjiang three strange, 21 Xi (醯), 22 1840, 23 Cu (醋), 24 180, 25 twenty-one days, 26 buy soy sauce, 27 deliciousness, 28 Qi Min Yao Shu, 29 Youshi, 30 parched rice color, 31 decoct the vinegar, 32 turn over the solid fermentation substrate of vinegar 33 saccharification, 34 be used as medicine, 35 gold medal award, 36 vinegar Dongpo, 37 Compendium of Materia Medica, 38 sterilization, 39 pottery jar, 40 Tao Hongjing, 41 rice bran, 42 grasp with hands, see with eyes, and smell with nose, 43 Chinese Medicine Dictionary, 44 Jia Sixie, 45 Health Care, 46 Bloody Fragrance 47 cooked glutinous rice, 48 Zhu Hengshun workshop, 49 Zhou Li, 50 bitter wine, 51 Daqu, 52 protection of national geographical indications, 53 appetizing, 54 killing evil poison, 55 Beigushan, 56 Jinshan, 57 Jiaoshan, 58 wheat, 59 wine with dregs, 60 Reeds cover VAT, Pole plus basket, 61 Xiangfei vinegar, 62 Shenxiu elder, 63 vinegar drying room.

To a certain extent, the result of word frequency can reflect the content of Zhenjiang vinegar culture, which has attracted more attention at the present stage, but there is a lack of connection between these entries. A single entry is not enough to form a whole story script for a picture book. It is necessary to sort these unrelated entries and find the connections between them so that they can be integrated into a story script.

2.3 Card Sorting

Card sorting is a way to make the organizational form of the information system meet the expectations of users, and information can be understood and classified through the user's construction of classification standards. There are two main types of card sorting, open card sorting and closed card sorting. Open card sorting does not provide the number and name of categories, which need to be customized by the participants, while closed card sorting provides the number and name of categories [11, 12]. This experiment uses the results of Zhenjiang vinegar culture based on big data statistics. The number of cards has been determined, and there are a large number of proper nouns. Therefore, closed card sorting is adopted, providing participants with the number and name of categories.

Hengshun

Hengshun: Jiangsu Hengshun Group Co., Ltd. (the original Zhenjiang Hengshun Sauce and Vinegar Factory) was founded in 1840 during the Daoguang period of the Qing Dynasty. It is the founder of Zhenjiang aromatic vinegar and a time-honored enterprise in China.

Fig. 1. An example of Zhenjiang vinegar culture card.

- 1) *Experimental materials*: The experimental materials include experimental instructions, card sorting tools (Usort & EZcalc), a classification package and 63 cards based on word frequency results after big data statistics. The card name is numbered 01–63 according to the word frequency. An example of cards is shown in “Fig. 1”.
- 2) *Participants*: A total of 20 participants from industrial design engineering, digital art, visual communication, computer technology, and other fields participated in the experiment, including 10 males and 10 females, with an average age of 27 years.
- 3) *Experimental process*: The material package with experimental instructions is sent to the participants, and the regulations of card sorting and experimental contents are introduced. In the process of card sorting, the participants are required to read the cards and the name of each category carefully, and then put each card into the most appropriate classification file folder.

3 Results and Discussion

3.1 Classification Results and Analysis

The data of each participant is imported into Usort and then analyzed by card analysis tool EZcalc. The Average Linkage method is used, and the correlation coefficient ranges from 0.3 to 0.7. In this study, 0.68 is selected as the standard for card sorting. A total of 63 Zhenjiang vinegar culture cards based on big data are divided into 7 categories, including process, efficacy, history, folk custom, brand, characteristic, and literature. The entry items of Zhenjiang vinegar culture corresponding to each category are shown in “Table 1”.

After classifying Zhenjiang vinegar culture entries based on big data, we need to analyze these categories and get the category entries with high word frequency and closely related to cultural phenomena. The total word frequency and the average word frequency of each category are calculated, as shown in “Table 2”.

According to the average value of word frequency, the four categories of brand, process, characteristic, and literature attract more attention, and the word frequency of “Hengshun” is the highest in the brand category. Hengshun Group, as the leading vinegar enterprise in Zhenjiang, has experienced hundreds of years of inheritance and development, and its development process can represent the typical vinegar culture of Zhenjiang. In the process category, the word frequency of “solid-state layered fermentation” is the highest. Zhenjiang vinegar is one of the typical representatives of solid-state layered fermentation process. Zhenjiang vinegar brewed by this process has the characteristics of “color, fragrance, thickness, mellowness, and acid”, and was included in the first batch of the National Intangible Cultural Heritage List in 2006 [13]. Among the

Table 1. CLASSIFICATION RESULTS OF ZHENJIANG VINEGAR CULTURE

No.	Category	Number of Vinegar Culture Entry
1	Process	03, 04, 07, 09, 10, 12, 16, 17, 25, 30, 31, 32, 33, 39, 41, 42, 47, 51, 58, 59, 60, 63
2	Efficacy	34, 38, 45, 53, 54
3	History	19, 21, 22, 23, 24, 28, 35, 37, 40, 43, 44, 48, 49, 50
4	Folk custom	20, 26
5	Brand	01, 11, 15, 52, 55, 56, 57, 61
6	Characteristic	02, 06, 13, 14, 18, 27
7	Literature	05, 08, 29, 36, 46, 62

Table 2. WORD FREQUENCY OF EACH CATEGORY

No.	Category	Total word frequency	Average Word frequency	High Frequency Entry
1	Process	978	44.45	Solid-state layered fermentation process
2	Efficacy	70	14.00	Be used in medicine
3	History	336	24.00	Vinegar protection regulations
4	Folk custom	72	36.00	Zhenjiang three strange
5	Brand	709	118.17	Hengshun
6	Characteristic	501	83.50	Fragrance
7	Literature	356	59.33	Heita invented vinegar

characteristic categories, the frequency of the word “fragrance” is the highest. As one of the Four Famous Vinegars in China, Zhenjiang vinegar has a unique and mild aroma that is different from other famous vinegars. In the literary category, the frequency of the word “Heita invented vinegar” is the highest. As a mythical story that took place in Zhenjiang, it reflects Zhenjiang vinegar brewing technique and the manual production method in the traditional society [14].

To sum up, after analyzing the word frequency results of Zhenjiang vinegar culture based on big data, four entries are extracted as the elements used in the design of Zhenjiang vinegar culture picture books. They are black tower vinegar, solid-state stratified fermentation craftsmanship, vinegar characteristics, Hengshun history.

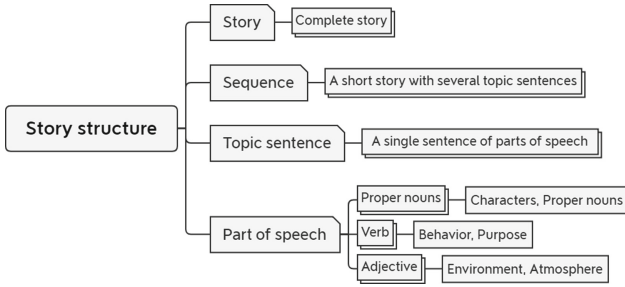


Fig. 2. Structure of the story.

3.2 Picture Book Story Construction

The picture book narrates the story through the combination of pictures and text, so the design of the picture book needs to have a complete story and a clear visual image. On the other hand, based on the traditional picture books, H5 digital picture books on mobile terminals can take advantage of H5 to digitize physical picture books, optimize graphics and text, increase multi-sensory channels, enhance audio-visual effects, and bring a good experience of reading to users [15].

After determining the cultural elements in the picture books, we need to integrate them into a complete narrative line. Therefore, it is necessary to create a story script structure that includes four cultural elements: Heita invented vinegar, solid-state layered fermentation process, vinegar characteristics, and Hengshun history.

Cheng introduced the structure of contemporary narrative theory by analyzing many French literary theorists. Among them, Todorov, a famous French literary theorist, established a narrative grammar covering all narrative works from the perspective of grammar theory in the Decameron. The narrative structure unit of a complete story is composed of story, sequence, topic sentence, and part of speech [16]. Its structure is shown in “Fig. 2”.

Story script writing is to integrate the extracted four Zhenjiang vinegar culture entries with the narrative structure of the complete story to express the cultural content of the four entries and construct a logical storyline at the same time. Among them, the entry “Heita invented vinegar”, as a literary legend introducing the origin history of Zhenjiang fragrant vinegar, has its own storyline. So, it is used as the story basis of scripts, while “solid-state layered fermentation process”, “vinegar characteristics” and “Hengshun history” are used as key clues to expand the story.

The script of this story is divided into nine sequences, numbered 01–09. Sequence 01–09 stories are as follows: (01) In the year when Du Kang invented the method of brewing wine, the whole family came to Zhenjiang. He opened a small workshop to produce and sell wine, and his son Heita helped him brew wine. (02) One day, after Heita finished his work and added a few buckets of water to the distiller’s grains, he went to do something else. A few days later, he picked up the wine jar, began to drink, and soon got drunk and fell asleep. (03) In his sleep, Heita dreamed of an old man with white hair. He told Heita that the water in the distillers’ grains has been fermented for 21 days and can be tasted today. (04) Heita woke up from the dream, walked into the workshop with a puzzled face, took a sip doubtfully, and immediately felt its fragrance

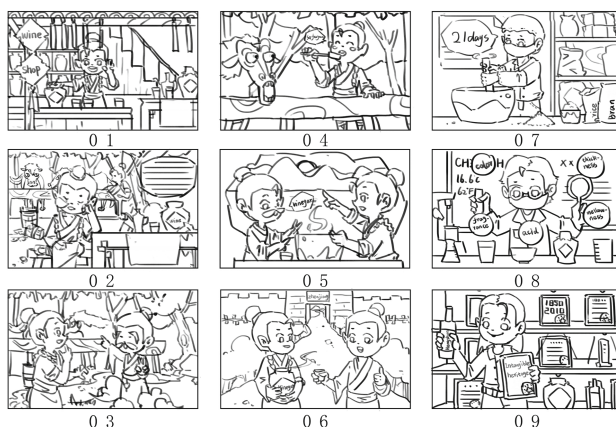


Fig. 3. Visual design of picture book of Zhenjiang vinegar culture.

and the taste of sour and sweet. (05) When Heita told his father what had happened in his dream, Du Kang wrote down the Chinese characters of 21 days (廿一) and Youshi (酉), thus a new Chinese character, “vinegar” (醋), was created. (06) When Du Kang and his son gave this seasoning to their neighbors, they all said it tasted good, and soon the vinegar became famous. (07) The legend that Du Kang brewed wine and his son invented vinegar has been handed down, and the people in Zhenjiang improved the process, by adding wheat bran, rice bran, and other raw materials and using a unique solid-state layered fermentation process to ferment the liquor. (08) After fermentation, Zhenjiang Hengshun fragrant vinegar has the characteristics of “color, fragrance, thickness, mellowness, and deliciousness”, and has become one of the Four Famous Vinegars in China. (09) This is the famous Hengshun fragrant vinegar. Hengshun began to add husk and rice bran to the distiller’s grains for solid-state fermentation as early as 1850. Due to this unique solid-state fermentation process, Hengshun fragrant vinegar brewing technique has been selected into the first batch of intangible cultural heritage and won many honors.

3.3 Construction of H5 Picture Book on Mobile Terminal

After completing the script of the picture book story, we designed the corresponding characters and the scenarios of each chapter according to the script. The characters and scenarios are mainly constructed according to the specific cultural content of the extracted four Zhenjiang vinegar culture entries and the plot of the picture book script. The scenarios of the picture book are drawn in the order of sequence number, as shown in “Fig. 3”.

After the story and vision in the picture book are determined, it needs to be integrated into the H5 picture book form suitable for the mobile terminal. In addition to the advantages of portability, the mobile terminal also largely relies on their hardware to improve the reading experience of H5 picture books, such as the application of sensors, cameras, and audio technology [17]. In this practice, different from traditional picture books, users can not only watch pictures and text, but also play audio to increase the



Fig. 4. Mobile terminal H5 picture book & QR code.

auditory experience. The user can select the appropriate background music when reading the picture book. A page of the H5 picture book is shown in “Fig. 4”, and you can read the whole book by scanning the QR code.

The above design practice shows that big data technology can help designers to obtain and analyze the content of Zhenjiang vinegar culture efficiently and conveniently in the case of the explosion of information resources and lack of professional knowledge. Taking Zhenjiang vinegar cultural innovation design as an example, the feasibility of applying big data technology to cultural innovation design is verified.

4 Conclusions

In the Internet era, information acquisition, reception, and processing are becoming faster and more convenient. Thanks to big data technology, we can quickly lock the target information and provide professional processing of massive information. Together with traditional methods of collecting and analyzing information, we can improve and learn from each other to make up for our weaknesses. In this study, we use big data technology to capture the cultural elements of Zhenjiang vinegar on the Internet. Combined with traditional research methods, the content of Zhenjiang vinegar culture can be analyzed in detail, objectively, and scientifically. On this basis, the cultural content of Zhenjiang vinegar is extracted through card sorting for the design of cultural communication in the form of H5 on the mobile terminal. It provides a new way to explore the innovative application of the same type of culture.

Acknowledgment. This work is supported by the University-Industry Collaborative Education Program, China (Grant No. 201902177003).

References

1. Z. Y. Shen, “Competitive advantage and development strategy of Zhenjiang balsamic vinegar,” *Jiangsu Flavoring and Non-ferrous Food*, vol. 22, no. 4, pp. 8-11, 2005.
2. W. L. Jin, “Information superhighway and cultural development,” *Social Sciences in China*, no. 1, pp. 4-15, 1997.
3. Z. Guo, “Culture: human procedure and human orientation,” *Journal of Renmin University of China*, no. 4, pp. 24-31, 2005.

4. L. W. Sun, G. H. He, and L. F. Wu, "Research on web crawler technology," *Computer Knowledge and Technology*, vol. 6, no. 15, pp. 4112-4115, 2010.
5. J. Yu and Q. Liu, "Review of topic web crawler research," *Computer Engineering and Science*, vol. 37, no. 2, pp. 231-237, 2015.
6. J. J. Zhao, S. L. Li, B. R. Sun, and M. H. Li, "A brief discussion on the network crawler technology based on python in the environment of big data," *China New Communications*, vol. 22 no. 4, pp. 68, 2020.
7. X. L. Du, "Vigilant literature research method," *Shanghai Education and Research*, no. 10, p. 1, 2013.
8. Y. Y. Li, "Field investigation: main research methods of cultural anthropology," *Today's Nationalities*, no. 9, pp. 45-49, 2007.
9. Y. Z. Zhu and J. Jing, "Research on Chinese word segmentation technology based on python language," *Communications Technology*, vol. 52, no. 7, pp. 1612-1619, 2019.
10. Mini Tag Cloud: Simple and powerful text cloud art generator, [online] Available: <https://www.weiciyun.com/>
11. J. Li, "Application of card classification in online game interface development," *Ergonomics*, vol. 17, no. 4, pp. 62-64, 2011.
12. N. N. Zhang, X. H. Liu, and H. Yuan, "Research on open card classification of digital service design technology," *Packaging Engineering*, vol. 39, no. 20, pp. 186-190, 2018.
13. Y. D. Zhu, X. B. Zou, J. Y. Shi, W. J. Zhao, and T. Z. Lin, "Analysis of Zhenjiang balsamic vinegar solid state fermentation process," *Chinese Journal of Food Science and Technology*, vol. 14, no. 8, pp. 256-261, 2014.
14. H. L. Wang, "Social model of technology," Nankai University, 2004.
15. Q. Y. Wu, "Development characteristics of H5 digital picture books in new media era," *Journal of Heihe University*, vol. 10, no. 7, pp. 193-194+203, 2019.
16. X. L. Cheng, "An overview of narrative theory," *Foreign Language Studies*, no. 3, pp. 10-15, 2002.
17. W. Zhang and H. T. Zhang, "Research on interaction design of H5 communication form on mobile terminal," *Journal of Hubei Institute of Fine Arts*, no. 4, pp. 117-120, 2018.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

