Universe Scientific Publishing

A Text Visualization Method Based on a Label Cloud

Mingao Fan, Xiaofei Li

Jilin Institute of Architecture and Technology, Jilin 130000, Changchun, China.

Abstract : An important direction of visualization technology research is the visualization of text data. Based on the characteristics of text information visualization, a text visualization method based on label cloud is studied, which puts forward the data index, complexity index and identification index to describe visualization, and calculates the weight of the total evaluation score by the calculation formula of three kinds of indexes. Through the visualization experiments of various kinds of text information, the results show that the method has some validity in visual measurement, and the index values at all levels are also relevant.

Keywords: Visualization; Label Cloud; Text

Introduction

Visualization is the theory, method and technique of using computer graphics and image processing technology to transform data into graphics or images to be displayed on the screen and interact with each other. It involves computer graphics, image processing, computer vision, computer-aided design and other fields, and becomes a comprehensive technology to study a series of problems such as data representation, data processing, decision analysis, etc. The virtual reality technology which is developing rapidly is also based on the visualization technology of graphic image.

The main forms of data storage and data propagation include text, and an important direction of visualization technology research is the visualization of text data. At present, the researchers put forward some methods of visualization technology research, such as text semantic structure method, label cloud method and so on [2-4]. A label cloud is a set of related tags and the corresponding weights. Weights affect the font size or other visual effects used. Tag clouds represent more power, and tags are typical hyperlinks that allow users to get a closer look at their content [5-6].

At present, the evaluation method of visualization technology is still in the development stage. The main work in the course of this project is to establish the measure of text data visualization method, calculate the weight value according to the calculation results, and analyze the application effect of text information visualization.

1. Text visualization method based on label cloud

The text visualization method based on label object is based on the establishment of three kinds of metrics, so as to calculate the weight according to the measure, and finally to evaluate the score calculation.

1.1 Data metrics

In visualization technology, the size of data is the basis of the entire quantification and evaluation process. Label clouds are presented from large to small according to the frequency of text data words, so the size index of data is an important indicator to evaluate visualization.

Suppose the collection of words in the text data is $S\{a_1,a_2,……a_n\}$， thereinto a_i (i= 1,2, , n) represents a word in a text message, makingN （a_i）Represents the number of words that appear in the text. Defines a collection of stop words that

[T] This work was supported by Jilin Province Innovative Training Program for college students (NO.202013604014)

appear multiple times in text data but have little effect on the content of the text.为 P={p_1,p_2, ……p_m}，For example Chinese appear in the words "yes", "one" and so on, in English, such as of, a, an, the and so on.

The steps for visualizing text data are：

① Filter the words in the text data - MMS words

② Text is filtered according to word collection S and stop word collection P to get S-P；

③ Calculate the frequency of words in S-P；

④ Select the frequency threshold h and filter the MMS word set C。

Based on the assumption, the total number of words in the text is $\mathrm{N}=\sum_{\mathrm{i}=1}^{m} m(a_i)$ ，The density of the word a in C is M(a)=m(a)/N，$a \in C$ ，The density of the words of pick-up is:

$$MI = \sum_{a\in C} M(a) = \sum_{a\in C} \frac{m(a)}{N} = \sum_{a\in C} \frac{m(a)}{\sum_{i=1}^{m} m(a_i)} \quad (1)$$

1.2 Complexity indicator

Complexity index mainly refers to the user's search and observation of text information, this method is mainly measured by direction measurement and letter-of-acceptance measurement.

MMS word measures refer to the aspect ratio of words in the visually determined area:

$$d = \max\{\frac{l}{w}, \frac{w}{l}\} \quad （2）$$

The upper type l is the length of the word picture and w is wide. The mean measure of all words in MMS Word Set C is:

$$D = \frac{d_1 + d_2 + \cdots\cdots d_n}{n} \quad （3）$$

The size of each word appearing in the visualization area depends on the weight, and the weight is significantly larger. But for some words with large length and small weight, MMS word measurements are also large. To solve this problem, increase directional measurements for evaluation. Suppose the angle of the word is u_i degree，The visualization determines the direction in which the area is represented as:

$$d(a_i) = \frac{\sum_{i=1}^{n} u_i}{n} \quad （4）$$

$d(a_i)$ The value range of is [0,1]，The larger the direction measurement, the higher the complexity metric.

1.3 Identification indicators

The identification index is mainly used to show the proportion of word color and the position composition of the visual display area.

Suppose that each text data display in the label cloud shows a different color, and the number of colors is n_c ，the number of MMS messages is n，The color weight is set to：

$$C = \frac{n}{n_c}(n \ge n_c) \quad （5）$$

There are often blank areas in the presentation areas of text visualization, which are measured by spatial utilization and can be increased by filling in the blank areas. Suppose the word occupies an area of area t， the area of the display area $t = W \times L$， Where W and L are the width and length of the display area respectively, the spatial utilization is：

$$T = \frac{t_1 + t_2 + \cdots\cdots t_n}{n} \qquad （6）$$

1.4 Weight calculation

Weights are calculated based on data size indicators, complexity indicators, and identification indicators. Weights are determined by fuzzy analysis. The above three indicators are compared between two and two, forming a fuzzy matrix B, and then turning it into a fuzzy consistency judgment matrix R：

$$B = \begin{pmatrix} 0.5 & 0.3 & 0.6 \\ 0.7 & 0.5 & 0.8 \\ 0.4 & 0.2 & 0.5 \end{pmatrix}_{3\times3} \quad R = \begin{pmatrix} 0.5 & 0.35 & 0.575 \\ 0.65 & 0.5 & 0.725 \\ 0.425 & 0.275 & 0.5 \end{pmatrix}_{3\times3}$$

The weight of the influence of data indicators, complexity indicators and identification indicators on the overall score is qw = (0.316,0.419,0.263).

1.5 Evaluation score calculation

According to the calculation and weight calculation of the corresponding index, the analysis of the text data information frequency algorithm is carried out, and the specific process of the algorithm is：

Step1: Text data information visualization parameters are initialized. For example, the maximum and minimum number of words displayed in the visualization determines the area, the maximum color of the display area, and so on.

Step2: Calculates the total number of words for text messages.

Step3: Determine the collection of MMS messages based on the filters and calculate the word information that needs to be visualized.

Step4: Initialize the visual area canvas and word information to display the words on the canvas.

Step5: The statistical words are calculated in area ratio and direction measurement. According to the calculation formula, the indicator value of the visual analysis is obtained.

2. Experiments and results analysis

In the experiment, through the network teaching students to the subject feedback subject information, using the label cloud visualization method for indicator calculation and analysis, so as to obtain the visualization results, as shown in Figure 1.

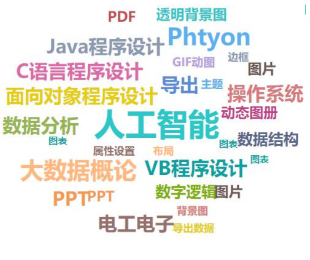

Figure 1 Text visualization results

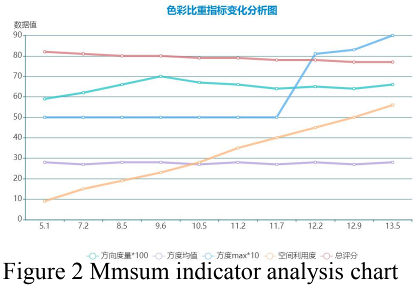

Figure 2 Mmsum indicator analysis chart

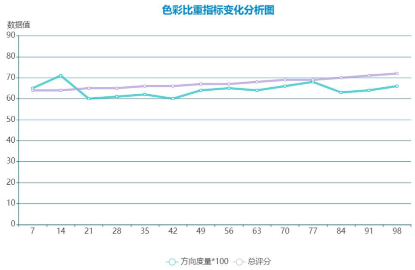

Figure 3 Color Gravity Indicator Analysis Chart

The direction measurement is calculated to be 0.61, MMS word density is 0.112, color gravity is 0.146, and space utilization is 1.689. The final total evaluation score according to the indicator is 65.12.

In the experiment, single text and multi-text are used to adjust the visual indicators in the algorithm and to analyze the results using multi-dimensional charts. The mmacance indicator line analysis graph is shown in Figure 2, and the color gravity analysis graph is shown in Figure 3.

As can be seen from the analysis chart, with the increase of MMS word density, the direction measurement is basically about 50 percent up and down fluctuations, the square max value gradually rises to a stable, the evaluation score gradually decreases, the visualization gradually deteriorates.

3. Conclusion

This paper gives a text visualization analysis method based on label cloud, mainly by establishing the text data visualization method measurement index, according to the calculation results of the weight value calculation, so as to analyze the application of text information visualization. Experiments show that the indicators are opposite to each other for text and multi-text information, and this method has some validity in visual measurement.

References

[1] Yang Xiaobo, Zhang Lin, etc. The design of the visualization of the data structure sorting algorithm, 2013(29):253-255.

[2] Ren Lei, Du I, Ma Shuai, et al. Overview of Big Data Visual Analysis. Journal of Software, 2014, 25(9): 1909-1936.

[3] Florian Heimerl, Steffen Lohmann, Simon Lange, et al. Word Cloud Explorer: Text Analytics based on Word Clouds[J]. IEEE Conference Publications,2014:1833-184.

[4] Jimmy Johansson, Camilla Forsell. Evaluation of Parallel Coordinates: Overview, Categorization and Guidelines for FutureResearch[J]. IEEE Transactions on Visualization and Computer Graphics, 2016, 22(1): 579-588.

[5] Yue Gang, Wang Nan. Study on the efficiency of knowledge visualization in e-learning. Software,2015, 36(2): 92-96.

[6] Rita Oliveira, Telmo Silva, Jorge Ferraz de Abreu. Development and evaluation of Clouds4All interface: A tag clouds reader for visually impaired users[J]. IEEE Conference Publications, 2015: 1-6.