The Design of Philosophy and Social Sciences Terms Dictionary System Based on Big Data Mining

Han-yang Li[1,2](✉)

[1] School of Humanities, Tongji University, Shanghai 200092, China
lihanyang1685@163.com
[2] School of Journalism and Communication, Minnan Normal University, Zhangzhou 363000, China

Abstract. Aiming at the problem that the traditional dictionary system cannot use the big data mining technology for term calculation, which leads to the long response time of the dictionary system retrieval, a philosophy and social science term dictionary system based on big data mining is designed. The hardware part designs the system controller and connects the single-chip microcomputer connection circuit. The software part first divides the term dimensions according to the characteristics of philosophy and social science terms, mines the corpus according to different dimensions, completes the calculation of philosophy and social science terms, sets up the term database structure, and finally completes the software design of the dictionary system. The experimental results show that: Compared with the traditional dictionary system, the search time of philosophy and Social Sciences terminology dictionary system based on big data mining is the shortest, and the detection accuracy of philosophy and social science terms is higher, which is suitable for all applications.

Keywords: Big data mining · Philosophy and social sciences · Term dictionary · Reaction time

1 Introduction

Glossary is an important tool for providing knowledge services in professional fields. However, there are still some problems in the compilation of existing term dictionary. For example, the knowledge content of terminology dictionaries is mostly simple, mainly providing explanations, English translation and other content, and the organization and description of deep knowledge needs to be improved. The degree of automation of term dictionary compilation is relatively low. Many term dictionary compilations still follow the traditional manual method. The process of term collection, collation, classification, typesetting, and proofreading is mainly done manually, and lacks the necessary automated auxiliary tools. These simple and repetitive manual labor are extremely error-prone and inefficient, leading to the compilation of term dictionary

S. Liu and L. Xia (Eds.): ADHIP 2020, LNICST 347, pp. 147–158, 2021.
https://doi.org/10.1007/978-3-030-67871-5_14

lagging behind the development of science and technology and the change of language facts, and it is difficult to achieve resource sharing [1]. How to deeply describe terminology knowledge from the perspective of knowledge organization, and then design a semi-automatic terminology dictionary compilation system, is an important topic in the current terminology dictionary research field and undoubtedly has very important significance.

Essentially, the compilation of terminology dictionary is an important part of knowledge production, and it is a frontier cross-cutting field of multiple disciplines such as lexicology, terminology, library and information science, computational linguistics [2]. The design of the terminology dictionary compilation system must first be based on knowledge organization, accurately reveal all kinds of knowledge behind the terminology, and form a unified and standardized knowledge representation framework. This requires the relevant achievements of lexicology, terminology and knowledge organization theory. Second, to achieve semi-automatic compilation of terminology dictionaries and improve the efficiency of knowledge production, it is necessary to actively absorb the achievements of computational linguistics in corpus construction, new word discovery, and terminology calculation. Finally, the terminology dictionary compilation has strong knowledge engineering features, and needs to realize the co-construction and sharing of knowledge, interactive collaboration and dynamic update from the perspective of project management.

The design of the term dictionary compilation system should rely on knowledge organization to form a more standardized and semi-automatic knowledge production process. The term dictionary is a tool that provides professional knowledge services, and needs to reveal in depth the objective things or knowledge content referred to by the term. Therefore, the compilation of term dictionary requires editors to have not only language knowledge, but more importantly, professional knowledge [3]. The term dictionary focuses on the concept of terms, and expresses these concepts in terms of words, which are generally sorted in order of topic. The conceptual category of terms and the relationship between category members are an important part of the term dictionary research. The term dictionary compilation is based on terminology and lexicography, applying basic methods and techniques of knowledge organization and computational linguistics. Norms, knowledge descriptions, knowledge links, etc., form a human-machine knowledge resource.

2 Hardware Design of Term Dictionary System

2.1 Design System Controller

The internal chip of the controller selects STC89C52 single-chip microcomputer as the processing core, and receives and processes the system-related index data [4]. Allocate MCU port to realize the control function of MCU. According to the I/O drive of MCU, the STC89C52 MCU has four groups of I/O ports P0, P1, P2 and P3. An external pull-up resistor is connected to the foot to ensure that the transistors in the internal output circuit of all ports are in the off state and the lower transistors are in the open state. Control the P0 port in parallel with two pull-up resistors to ensure the output of “0” and “1” processing instructions. The connection diagram of the single-chip microcomputer as the output port is shown in Fig. 1:

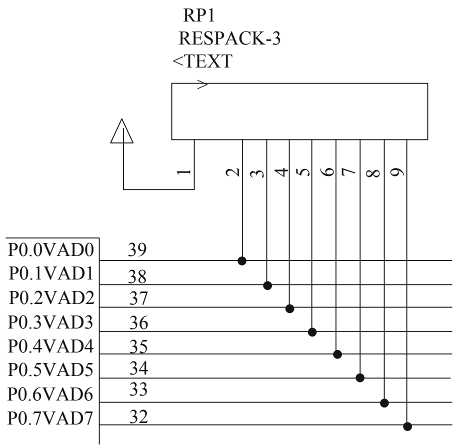

Fig. 1. Wiring diagram of the single-chip P0 group port used as output

As shown in Fig. 1 above, the P0 group port is connected to the circuit that drives the LCD display, the P1 group port is used to store philosophy and social science terminology data, the P2 group port controls the LCD display drive signal of the dictionary system, and the P3 group port is used The change of terminology and the detailed allocation of I/O ports are shown in Table 1:

Table 1. MCU I/O port allocation table

Serial number	I/O port	Allocation function
1	Group P0	Provide data signal for LCD display
2	P1.0 pin	Receive display output
3	P1.1 pin	Receive data signal output
4	P 1.2–p 1.4 pin	Provide driving signal for LCD display
5	P1.5 pin	Driver chip controller
6	Group P2	Drive control chip
7	P3.2–p3.4 pin	Receive key input
8	P3.5 pin	Receive controller output

According to the port function shown in Table 1, connect each port of the chip, design the control circuit of the controller, and complete the design of the hardware part [5].

2.2 SCM Connection Circuit

The system uses a single-chip microcomputer as the core processing unit, combined with resistors and capacitors and other devices, and uses the single-chip microcomputer as the smallest processing unit. The final control core circuit is composed of a block diagram, as shown in Fig. 2:

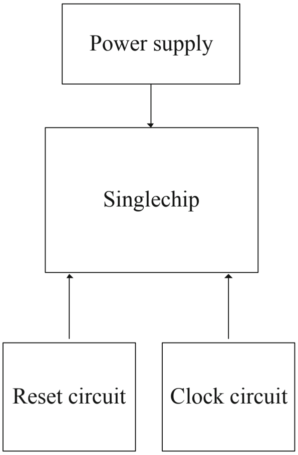

Fig. 2. The minimum circuit composition of the STC89C52 microcontroller

As shown in Fig. 2, the left side of the single-chip microcomputer is mainly connected to the power circuit, and the right side is connected to the clock circuit and the reset circuit. When the single-chip microcomputer is actually working, the time for the single-chip microcomputer to access the storage from the ROM is defined as a machine cycle. Store access data for one machine cycle [6]. Use the XTAL1 and XTAL2 ports of the oscillator as the oscillator input/output ports. The XTAL1 port of the oscillator uses an internal and external clock to connect a quartz crystal, and then an external capacitor is mounted to form a parallel resonant circuit to allow the internal oscillation circuit to generate self-oscillation (Fig. 3).

In order to prevent the SCM from being disturbed by the environment and causing the dictionary system to malfunction, the reset circuit of the SCM system is adjusted to a level switch reset method, so that the capacitor charge is in a short circuit state when

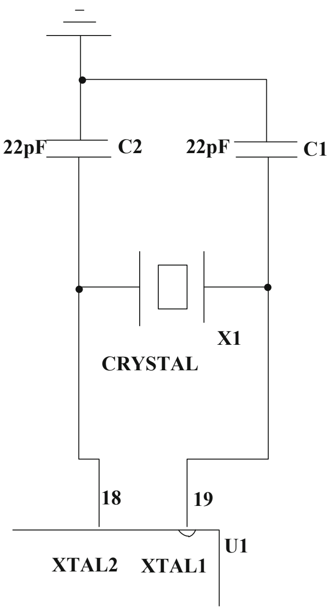

Fig. 3. MCU clock circuit diagram

the dictionary system is turned on, and the reset pin is adjusted Connect to high level [7]. After the power supply is stable, the reset pin is grounded through the resistor, so that the capacitor plays the role of isolating the DC level. Design the reset circuit of the single chip microcomputer as shown in Fig. 4:

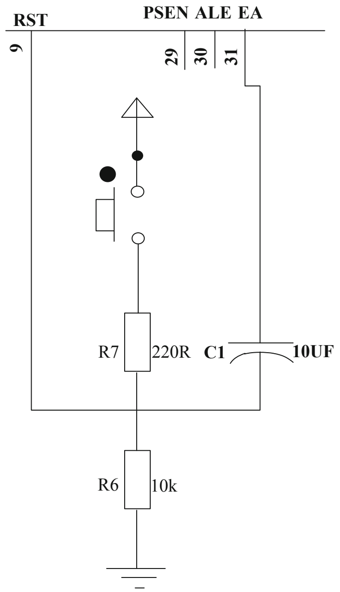

Fig. 4. Reset circuit diagram of single chip microcomputer

The power circuit part adopts USB to directly provide 20 V DC power. In order to ensure that the reading of the program starts from internal storage, the EA pin in the above figure is connected to a high level to complete the hardware design of the dictionary system.

3 Software Design of Term Dictionary System

3.1 Use Big Data Mining for Term Calculation

When using big data mining and terminology calculation, according to the characteristics of philosophy and social science terms, according to the following process of mining, the mining process is shown in Fig. 5:

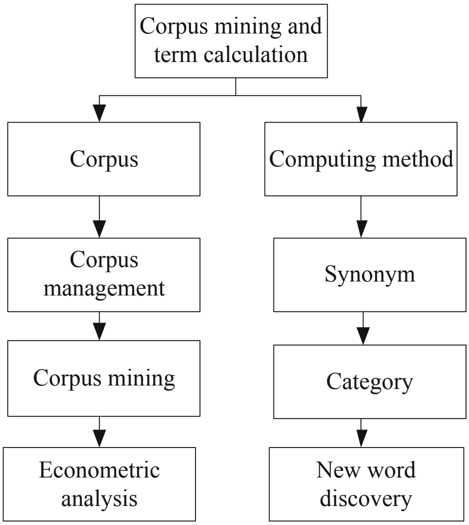

Fig. 5. Terminology mining process

According to the mining process shown in Fig. 5, the philosophical corpus is first divided into four dimensions according to the different learning styles, sensory/intuitive, visual/linguistic, positive/reflective and sequential, as shown in Table 2:

Table 2. The four dimensions of philosophy and social science terms

Dimension	Classification of learning styles	Source of corpus
Dimension 1	Active type	Philosophy innovation theory
	Reflective type	Problem
Dimension 2	Language type	Literal interpretation
	Visual type	Media publicity
Dimension 3	Sensory type	Examples, facts
	Intuition type	Overview
Dimension 4	Global type	Academic research
	Sequential type	Material science

According to the four dimensions shown in Table 2, each dimension corresponds to the source of philosophical and social science corpus of different dimensions, mining the corpus according to different dimensions, summarizing all the above corpus data, and describing the characteristics of social science terms using Jones matrix Words:

$$M_j(w) = e^{\beta(w)} U(w) \tag{1}$$

In the above formula, β indicates the amount of terms that do not affect the description mining, w indicates the frequency of each material component relative to the term, U indicates the frequency function, and j indicates the number of types of terms. To process the philosophical corpus data summarized above, assuming that at this time R_{in} represents all philosophical material data, and R_{out} represents output philosophical social science terms, then the two processing processes can be expressed as:

$$\begin{cases} R_{in} = \begin{pmatrix} e^{-\varphi} \cos \xi & -e \sin \xi \\ e^{-\varphi} \sin \xi & e \cos \xi \end{pmatrix} \\ R_{out} = \begin{pmatrix} e^{-\psi} \cos \xi & -e \sin \xi \\ e^{-\psi} \sin \xi & e \cos \xi \end{pmatrix} \end{cases} \tag{2}$$

In the above formula, (φ, ξ) and (ψ, ξ) are used to describe the input and output of the philosophy and social science PSP, and the final output of the above formula is R_{out} philosophy and social science terms. Due to the repetitiveness of the mining objects, the excavated terminology is synonymous. Use the obtained PSP to determine synonyms. The calculation formula is:

$$M_d(\Delta t) = \begin{pmatrix} e^{jw} & 0 \\ 0 & e^{-jw} \end{pmatrix} \tag{3}$$

In the above formula, $M_j(w)$ represents the scope of philosophy and social science terms, so for any philosophical literature, the discovery process of philosophy and social science terms can be expressed as:

$$\overline{J}_{out} = M_j(w)R_{in} \tag{4}$$

In the above formula, $\overline{J}_{out}$ represents the output of philosophical and social science terms. After using big data mining and terminology calculation, the calculated philosophy and social science terms are aggregated into a dictionary, and then the dictionary is converted into a philosophy and social science terminology database, and the software design of the philosophy and social science term dictionary is finally completed [8].

3.2 Design Terminology Database

Before designing the terminology database, set up the structure of the terminology database according to the philosophical terminology data dictionary obtained from the above calculation. The designed database structure is shown in Table 3:

Table 3. Designed database structure

Terminology name	Identifier	Type	Null value
Dialectics	id	varchar	no
Whiteboard	openid	varchar	can
Superman	opendate	varchar	no
Transcendental	applicationid	varchar	can
Contemplation	applicationdate	varchar	can
The world of existence	maingenusid	varchar	can
List	genusid	char	no
Dionysian	applicationcontryid	varchar	can
Public will	contryid	char	can
Leap of Faith	patenttype	varchar	no
Heisenberg Uncertainty Principle	name	char	no
Generated world	claim	char	no
Pascal's bet	interapplication	varchar	can
True self	interopen	char	can
Thinking self	incomedate	varchar	no

According to the database structure shown in Table 3, the philosophy and social science terminology is converted into different social science categories, and this database structure is managed using the existing management form of the computer in order to realize the retrieval function of the philosophy and social science terminology system [9, 10]. The database index is not designed for full-text indexing. When using keywords to query, the database search process becomes a traversal process similar to page-by-page flipping. For an efficient retrieval system, the key is to establish a reverse indexing mechanism similar to the scientific and technological index. When storing the data source, such as a series of articles, in sorted order, there is another sorted keyword list.Used to store the mapping relationship between keywords and articles, the retrieval process is the process of turning fuzzy queries into a logical combination of multiple precise queries that can use the index [11, 12]. As a result, the efficiency of multi-keyword query is greatly improved. Therefore, the term retrieval problem is ultimately a structural ranking problem. Therefore, Lucene is used to redefine the terminology database composition structure during retrieval. The defined composition structure is shown in Table 4:

Table 4. Database search composition structure

Serial number	Composition structure	Search function
1	Org. apache. Lucene. search/	Search entrance
2	Org. apache. Lucene. index/	Search entrance
3	Org. apache. Lucene. analysis/	Search entrance
4	Org. apache. Lucene. queryParser/	Search entrance
5	Org. apache. Lucene. document/	Search entrance
6	Org. apache. Lucene. store/	Underlying IO storage structure
7	Org. apache. Lucene. util/	Some common data structures

According to the database search composition structure as defined in Table 4, the retrieval of the philosophy and social science terminology database is finally realized, and the software design of the philosophy and social science terminology dictionary system is completed.

4 Simulation Experiment

4.1 Build System Experiment Framework

When constructing the experimental system framework, enter keywords in the system to search and query according to the requirements of the dictionary system, give relevant explanations, the server performs the search operation after receiving the searched query, and returns the search results and displays to Queryer, so the experimental framework of the system can be built as:

According to the system experiment framework shown in Fig. 6, two traditional philosophical and social science term dictionary systems and a big data mining-based philosophical and social science term dictionary system are used to conduct experiments, and three systems are set to query 100 to 500 philosophies respectively. Social science terms, comparing the reaction time of three design systems to different amounts of philosophical terms.

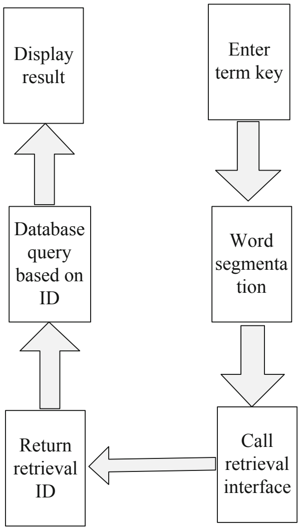

Fig. 6. The system experiment framework built

4.2 Analysis of Experimental Results

Based on the above experimental preparation, the system's detection program is called, and the received data packets of the three systems are detected to be in the normal receiving state. The definition begins to be retrieved. The interface displays the term interpretation as a reaction time. The query time fed back by the dictionary system under the number, and the final experimental results of the five systems, are shown in Table 5:

Table 5. Experimental results of three dictionary systems

Number of terms	Average time of system response (ms)		
	Traditional dictionary system 1	Traditional dictionary system 2	Designed dictionary system
100	10.54	8.64	3.58
200	9.65	6.57	3.62
300	10.20	7.23	2.68
400	10.68	8.32	3.57
500	9.88	6.48	3.84

During the experiment, the debugging program is called to detect the reaction time of the system to a philosophical and social science term. From the experimental results, it can be seen that the traditional dictionary system 1 has the longest response time to a term retrieval among the three dictionary systems, and the system's timely response is weak Compared with traditional dictionary system 1, the traditional dictionary system 2 has a shorter retrieval response time than the traditional dictionary system 1, but there is still a delay. The average retrieval response time of the philosophy and social science term dictionary system based on big data mining is faster than the two traditional ones. System, the timeliness of the system is strong, and it is more suitable for practical application.

On this basis, the accuracy of the three systems for the detection of philosophy and Social Sciences terms is analyzed, and the comparison results are shown in Fig. 7.

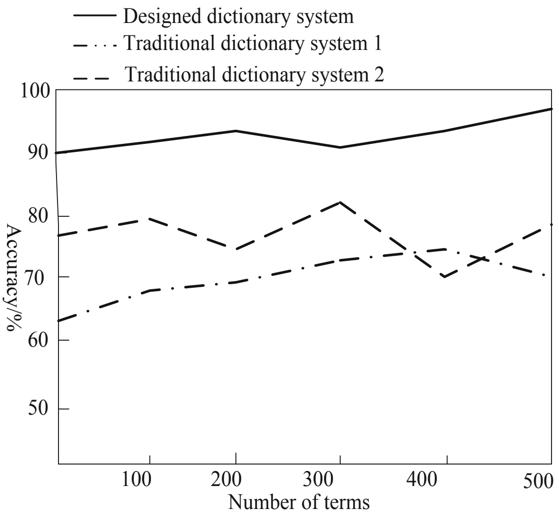

Fig. 7. Accuracy test in terms of philosophy and Social Sciences

It can be seen from Fig. 7 that the average retrieval accuracy of traditional dictionary system 1 for 500 words is 68%, that of traditional dictionary system 2 for 500 words is 74%, while that of philosophy and Social Sciences terminology dictionary system based on big data mining is 94% on average. Thus, it can be seen that the philosophical and social science term dictionary system based on big data mining has a positive impact on philosophy The accuracy of social science terminology detection is high.

5 Conclusion

The term dictionary compilation requires a more general knowledge organization model to provide a framework for the design of dictionary compilation systems. Furthermore, the concepts of user interaction, dynamic update, and term calculation in knowledge organization research are introduced into the dictionary compilation process, and a semi-automatic term dictionary auxiliary compilation system is designed.

This design properly integrates process management, term calculation, user interaction, etc., and helps to improve the quality and efficiency of term dictionary compilation. Using the existing professional literature database as a rough corpus, it is convenient for compilers to choose vocabulary, quantitative analysis and knowledge extraction, and improve work efficiency. The dictionary data is multi-dimensionally linked according to the semantic structure of knowledge organization to form a multimedia representation, which helps users understand the relationship between different concepts and improve the efficiency of knowledge learning. Strengthening the research on terminology calculation, scientific and technological corpus construction, and the formation of term-oriented automatic processing methods and technologies is a subject that needs to be further enhanced.

References

1. Wagreich, M., Sames, B., Hu, X.: Climate-environmental deteriorations in a greenhouse earth system: causes and consequences of short-term cretaceous sea-level changes (a Report on IGCP 609). Acta GeologicaSinica (Engl. Ed.) **93**(S1), 144–146 (2019)
2. Jia, P., Xue, H., Liu, S., et al.: Opportunities and challenges of using big data for global health. Sci. Bull. **64**(22), 1652–1654 (2019)
3. Susan Smith, N.A.S.H.: The role of big data and machine learning in the integration and implementation of historical, current, and continuously gathered earth data. Acta GeologicaSinica (Engl. Ed.) **93**(S1), 56–58 (2019)
4. Zheng, P., Shuai, L., Arun, S., Khan, M.: Visual attention feature (VAF): a novel strategy for visual tracking based on cloud platform in intelligent surveillance systems. J. Parallel Distrib. Comput. **120**, 182–194 (2018)
5. Boldosova, V.: Deliberate storytelling in big data analytics adoption. Inf. Syst. J. **29**(6), 1126–1152 (2019)
6. Zhao, Y., Zhong, H., Xu, S., et al.: Quantitative expression of paleogeographic information based on big data. Acta GeologicaSinica (English Edition) **93**(S1), 83–85 (2019)
7. Li, M., Fu, J., Chen, A., et al.: The connation of geological big data mining based on cloud computing. China Min. Mag. **28**(S1), 343–346+348 (2019)
8. Zhao, Y., Zhang, Y.: Design of text information mining system in big data environment. Mod. Electron. Tech. **41**(1), 125–128 (2018)
9. Chen, B., Sun, L., Han, X., et al.: A bridge-based lexicon learning method for semantic parsing. J. Chin. Inf. Process. **33**(5), 24–30 (2019)
10. Bao, H., Li, L., Xu, W., et al.: Data mining of psoriasis prescription for external use: based on dictionary of traditional Chinese medicine prescriptions. J. Tradit. Chin. Med. **60**(11), 974–978 (2019)
11. Shuai, L., Gelan, Y.: Advanced Hybrid Information Processing, pp. 1–594. Springer, Heidelberg
12. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. Front. Genet. **10**, 120 (2019)