Research on the development of scientific and technological intelligence in big data environment

Qiong Wu[1,* a], *Chensheng* Wu[1,b], and *Yanjun* Liu[1,c]

[1]No.140, Xizhimenwai Street, Xicheng District, Beijing, 100044 P R CHINA.

[a]wwqqppdd@163.com, [b]wu1082@163.com, [c]65415643@qq.com

Abstract. This paper describes the features of Internet data from the perspective of the development of scientific and technological intelligence in the era of big data. It has discussed the main problems that affect the construction of scientific and technological intelligence system and the specific technical problems in the process of construction. What's more, it has clarified that in the big data environment, the basic features of the development of scientific and technological intelligence and problems faced by the big data intelligence system. Finally, the author has predicted the opportunities and challenges for scientific and technological intelligence service agency posed by the big data.

1 Basic features of the big data environment

The scientific and technological intelligence has gone through the era of intelligence scarcity and intelligence popularization, and is moving towards the age of intelligence holography. Only the big data can complete the holographic description of the service object. Now the big data era has gradually shifted to the "artificial intelligence era," and also has made a profound impact on the enormous changes brought about by the external environment in the field of scientific and technological intelligence. For scientific and technological intelligence, how much data is called big data? It does not really matter whether the measurement standard is in TB, ZB, or PB, and what's important is that we can see the panorama of the intelligence service, also known as holography. For this reason, we have to change the way of thinking about data processing, and the mistaken concept of big data that its data and data structure obtained by technological means [1].

1.1 Distributed data and un-structured data

The data is usually divided into two categories in the industry: structured data and unstructured data. However, in the context of big data, the traditional means of processing structured data has all but failed due to the millions of data sources. As a result, we have to process the structured data we search by using the means of processing unstructured data.

---

* Corresponding author: wwqqppdd@163.com

That is, when we process hundreds of structured data sources, we can process them according to their structural features. However, if the quantity increases, thousands or even tens of thousands of disorder structured data cannot be processed according to their structural features and must be dealt with by unstructured means.

1.2 The Internet is big and comprehensive "database"

There is no denying the fact that the Internet is the largest and most comprehensive database. In the traditional sense, the scientific and technological community is desired to find the largest and most authoritative database, thus obtaining the most needed information content. With the development of the Internet, the traditional structured database content can often be retrieved on the Internet[2]. However, the information on the Internet cannot be compared with the structured data that has gone through secondary processing in terms of quantity, coverage, comprehensiveness and speed of response.

2 Core issues of big data intelligence system

The core issues of big data scientific and technological intelligence system are the intelligence search capabilities and human reading ability. Identifying and obtaining the vast amount of information (also known as big data) from the largest "database" - the Internet – serve as the foundation of big data scientific and technological intelligence system. How to convert big data information into information with tens of thousands of words that are suitable for people to read is a problem that must be solved in processing big data.

2.1 basic features of scientific and technological intelligence system

The intelligence collection of the intelligence system requires All-Source Intelligence, which requires collecting all information sources on the Internet and multivariate & heterogeneous data from toll database and internal enterprise information, which is the reason of creating big data. In fact, only big data can fully reflect the full range of intelligence objects, so we must make network-wide collection when we collect information. Otherwise, it is very difficult to collect the order parameter that reflects the essential changes of intelligence objects. According to the theory of H.Haken, the order parameters are the decisive factors in the transformation of old and new systems, where the order parameters will change. Therefore, the so-called intelligence system tracking only a few dozen or hundreds of websites cannot actually track the key order parameters of the intelligence system, and is also difficult to give a prediction and warning of the revolutionary system changes with the greatest intelligence value. This is one of the most urgent tasks that the construction is actually going to accomplish.

2.2 The core driving force of scientific and technological intelligence system under big data

The capability of identifying and judging demand and applying big data processing systems to identify data associations and making judgments is a big change in the importance of scientific and technological intelligence in the era of big data. First, intelligence points' position discrimination capability in the global. This will answer whether all your intelligence information covers all of the key points (or points of knowledge) in the field, and coverage of manually processed intelligence in the global. Second, classification

capability. The classification of the search information through mathematics method, models and tools is essential in the era of big data. This capability has turned from general capability to one of the core capabilities in the age of manual processing data. Third, concentrating capability. The concentrating capability that can concentrate big data with little intelligence information into "high concentration" information and match with human reading ability serves as the foundation of human reading ability. Information visualization is also an important step in human reading ability. Fourth, associating capability. The basic feature of big data is data association, but how to identify the data association and its logical relationship through mathematical, psychology, sociology and other means is the capability that the scientific and technological intelligence should identify beyond IT's data association. Fifth, interpretation auxiliary capability. Intelligence interpretation is the core content of intelligence. The intelligence interpretation is traditionally based on personal knowledge, perception, sensitivity on the information and data. In the era of big data, the interpretation of information data must be related to the data. While interpreting the data information, the auxiliary system provides the related knowledge of data information, i.e. the related links, which will help the readers (usually technical experts) make more comprehensive and profound understanding of the content behind the information data, inspire and assist readers to understand the objects, so that they can make a more comprehensive and systematic interpretation, which is also known as "interpretation supporting system or identification supporting system." Sixth, report assisted generation. After interpreting scientific and technological intelligence, it is usually necessary to provide a consultation report with the demander. The writing of the report often occupies a large amount of time of the researchers. In fact, this report requires the scholars to re-summarize the knowledge in relevant fields[3]. The big data intelligence system can provide similar contents that are contained in the big data based on experts' outline or paradigm, and then the experts can modify them accordingly, thus quickly completing the scientific consultation report. Seventh, capability of identifying industry experts. Scientific and technological intelligence is interpreted and determined by the human brain, and possessor of domain knowledge – experts, how to choose them is the key to the interpretation of scientific and technological intelligence. In order to find the competent experts to perform right thing, intelligence agencies required to establish a comprehensive network that contains huge expert information, structure and relationships.

2.3 Intelligence system integrated workshop hall's core capability feature

With the model that is built based on the mathematic, social simulation knowledge, it can use the "science" mode of thinking to help solve the qualitative problems of "liberal arts", and quantitatively give accurate explanation, and achieve changes from quantity to quality on this basis. In fact, this is the concept of the integrated workshop hall proposed by Mr. Qian Xuesen in the early nineties. This concept has laid a solid theoretical foundation for the intelligence community to meet the era of big data and also has pointed out the direction for the construction of scientific and technological intelligence system in the era of artificial intelligence. On this basis, the mathematical model analysis of practical engineering is of great significance in the quantitative analysis under qualitative guidance. On the basis of this engineering model and tool, the identification of correlation supply of knowledge and opinion behind it and digging of more valuable information from the data are the correct understanding of intelligence interpretation. The development process from factual data to process-based data is a change based on the understanding of the data, which has changed the data from "dead" to "live" and changed from the happened facts to the ongoing facts. That is: with various means such as data mining in the IT industry, under the guidance of

the system of science, it will identify its association and laws, and help make the intelligent, scientific decision, which is the basic path for the development of big data.

3 Problems faced by scientific and technological intelligence system in the big data environment

3.1 Problems faced by professional intelligence agencies

In order to truly and clearly understand the needs of scientific and technological intelligence, and interpret the intelligence needs with more accurate language, the professional intelligence agencies should possess over 10 years of intelligence research experience and obtain background knowledge in terms of IT and mathematical modeling, thus jointly completing the research and development of intelligence system. Since the intelligence is a discipline that studies complex systems, the development of information systems cannot solely rely on computer technicians, which requires systematic and scientific thought, foundations of mathematical modeling, concepts of complex social systems, awareness of social sciences and long-term experience. It is precisely because of this, Professor Shen Guchao of Nanjing University has put forward: the academic circles feature lack of verification and poor operability due to the emphasis on analysis of various models of theory, methods and tools; business circles feature strong universality, insufficient individualization, different from the description of academia and failing to meet demand for intelligence in the market competition due to the perspective of quantitative analysis of business intelligence; the industry circle accepts the concept proposed by the academic circles and uses the information system developed by the "business circle", which has caused many questions in the big data products[4].

3.2 Dilemma of professional and technical layer

First, capacity of the database. At present, the "small" common database can only accommodate millions of records, and tens of millions, hundreds of millions of data will encounter when we operate big data system. The "big" database is inconvenient to use. Second, Windows program installation. Big data processing system will operate dozens of applications simultaneously. However, Windows will drastically reduce its operational efficiency after installing multiple software programs, and even cannot function properly. Sometimes, the system will crash. Third, disk file format. The file format in the original Windows XP only supports 2G, which has made restrictions for the big data files with over 10 G. Other problems exist in the Win7 operating system. Excel2003 can only process 60,000 rows of records, while Excel 2007 and 2010 can only process one million rows of records, which have not taken the big data demand into consideration. Fourth, I/O capacity. When processing big data, the cache of disk system actually does not work, and the disk reads and writes data at mechanical speed, which is one millionth of processing speed of CPU cache and other electronic transmission. It has formed a bottleneck in the big data processing. Even Hadoop cannot solve the problem, unless large-scale data center can provide services at any time. The best way to process I/O at this point is the SSDs, the solution of using laptops + SSDs can perform better than that of data center + disk arrays. Fifth, memory problem. No matter how much memory windows can manage, they cannot keep up with the expected memory for big data processing. Although Hadoop can manage larger memory, it cannot solve the problem of transferring data between memories in different machine (the transmission speed of the network cable is one thousandth of the memory speed). The operating system of parallel computing system has not fully supported

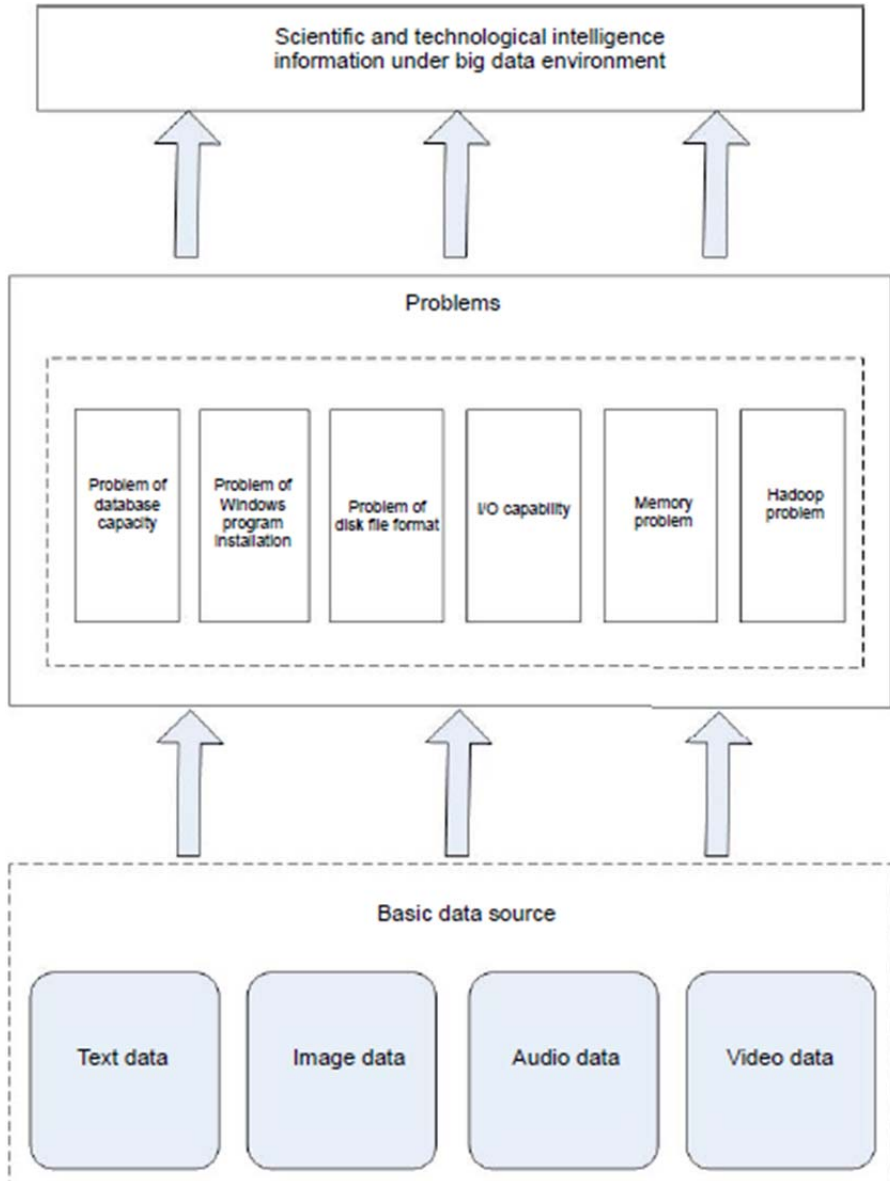the big data system. Sixth, Hadoop problem. The parallel processing cannot be cross-calculated[5],see Figure 1.

Fig. 1. Problems of professional and technical layer

At present, in the context of artificial intelligence, the IT industry is still exploring the true meaning of big data. For example: the debate on unstructured databases, C++ and Java are suitable for processing big data, but the scientific and technological intelligence industry has changed from the urgent desire to obtain intelligence information in big data to instruct service objects "what to do and how to do". Therefore, the current intelligence industry should use existing IT means to "summarize" large amount of disorganized original "data", then extract the content with core values and organize them into the content that the demand side strongly needs, thus extracting valuable core information. On this

basis, it will complete the extraction of relevant intelligence information, apply in practice, and ultimately complete the countermeasures and suggestions report under the support of big data, which is the framework of the intelligence system services in big data environment.

4 Opportunities and challenges for scientific and technological intelligence service agencies presented by big data

The storage, analysis and processing of intelligence information in the era of big data is an important means of acquiring knowledge and also the key content of intelligence research. With the rapid development of digital technology and network, massive unstructured and rapidly growing data continue to emerge, causing a great impact on intelligence research. Intelligence studies will inevitably face a technological revolution. From the date of the birth of scientific and technological intelligence, it has never been a comprehensive grasp of intelligence objects. This possibility is realized for the first time in the context of the big data era. The scientific and technological intelligence that historically relies on the exclusive source of information as the feature, the ability to interpret intelligence as the core competitiveness, has developed and enabled the scientific and technological agencies to empower unique ability to serve through the application of unique big data tools as a means, safeguarding the interpretation of intelligence as the goal and using semi-automatic intelligence as the core capability under the guidance of unique intelligence thinking. The big data helps scientific intelligence agencies accomplish its mission and provides the theoretical and practical basis of identifying differences and existence between intelligence agencies and general soft science agencies [6] .

References

1. Herman Haken [M]. Germany: Shanghai Translation Publishing House, 2001.
2. Qian Xuesen, Yu Jingyuan, Dai Ruwei. A New Scientific Field - An Open and Complicated Giant System (Including Social Systems) and Its Methodology [J] .Journal of Nature, 1990 (1).
3. Qian Xuesen, Open Complex Giant System [J]. Pattern Recognition and Artificial Intelligence, 1991 (1)
4. Dai Ru-wei. Metasynthesis from qualitative to quantitative [J]. Pattern Recognition and Artificial Intelligence, 1991 (1).
5. Dai Ruwei. Metasynthesis Method from qualitative to quantitative (Met synthesis) - Open complex giant system methodology [M] Beijing: Tsinghua University Press, 1996.
6. Dai Ruwei. The Proposal and Recent Development of Metasynthetic Method (M) from Qualitative to Quantitative [J]. Natural Science, 2009, 31 (6).