Innovation Research of Scientific and Technical Information in the Age of Big Data

Liu Ru [1, a *], Li Rong [1, b] and Wu Yuhui [1, c]

[1] Beijing Municipal Institute of Science and Technology Information, China

[a]rudyliu1983@hotmail.com, [b]lirong_dragon@163.com, [c]wyh-6969@163.com

Keywords: Big Data; Scientific and Technical Information; Innovations

Abstract. In the age of big data, one of an important means to develop industry, science and technology in China is to explore the dynamic information solution based on big data. With popularization of smart phones and mobile terminals and wide coverage of networks of 3G/4G/WIFI, the data related to every one of us will become the data to be reported and analyzed; this is a new chance and challenge of scientific and technical information and an inevitable new environment to it. The age of issuing, sharing and applying data massively has started a reform of scientific and technical information.

Introduction

With development of globalized economy, a larger application market will appear in China, which makes China one of the countries with the biggest data. The age of big data is an inevitable result of the development of network technology and improvement of massive data processing capacity. In February 2013, Ministry of Industry and Information (MII) issued Guidance about Data Center Construction and Structure [1], where the basic principles of construction and structure of data base are determined, including principle of market demand orientation, principle of resource environment priority, principle of regional planning and coordination, principle of considering multiple factors and principle of focusing on both development and safety. It suggests that the government has implemented the measures of facing the age of big data and has made full preparation in terms of both policy support and financial support. It's shown in Prediction and Analysis 2012-1016 of Big Data Technology and Service Market in China, the first report about the technology and service market of big data in China issued by International Data Company (IDC) [2], that the market scale will increase from 77.6 million dollars in 2011 to 617 million dollars in 2016. The compound growth rate in the next five years will be 51.4% and the market scale will increase for 7 times. The income of global market of big data technology and market in 2016 will be 23.8 billion dollars, nearly 150 billion RMB Yuan. The large international companies, including IBM, Google, Amazon and Microsoft, all have set foot into the market of big data to earn their share.

Therefore, in the age of big data, one of an important means to develop industry, science and technology in China is to explore the dynamic information solution based on big data. With popularization of smart phones and mobile terminals and wide coverage of networks of 3G/4G/WIFI, the data related to every one of us will become the data to be reported and analyzed; this is a new chance and challenge of scientific and technical information and an inevitable new environment to it. The age of issuing, sharing and applying data massively has started a reform of scientific and technical information.

The concept of "big data" for scientific and technical information

It should be pointed out that the concept of big data has been discussed hotly in recent years, but what on earth is the big data? Many people are still not clear about it. Firstly, it's sure that the experts in the industry generally agree that "big data is not just about more data". The most popular two concepts are:

Concept 1: in 2011, Doug Laney pointed out the "3V" features of big data: Volume, Velocity and Variety. After that, the feature was expanded to 11V, including Validity, Veridicality, Value and Visibility, etc.

Concept 2: Big data is a type of technology and was pointed out in as early as 2001. At present, the popularity of big data is not just more Volume, Velocity and Variety than a decade ago but the evolvement of data processing capacity, innovation of data processing means, etc. driven by the new technologies. In other words, the recent big data is the data ignored for the limitation of technology. It can be seen from the description of the concept of big data by various industries that the basis of the big data is the significant increase of data and mature technology (see the figure below).

Figure 1. The Essence of Formation of Big Data

Firstly, the data are increasing dramatically. With popularity of mobile terminal and massive coverage of internet and rise of social network data, search data, online transaction data and data of Internet of Things, the basis is also laid for big data analyzing; secondly, the technologies related to big data are getting mature, including datamation technology, capacity of storing and processing massive data, capacity of analyzing the massive and real-time data. These technologies have made amazing progress in recent years, which plays an essential role to witness the arrival of the age of big data.

Although the concept of "big data" hasn't been finally defined, in terms of scientific and technical information, the purpose of big data lies in exploring valuable scientific and technical information from the massive structured and unstructured data for relevant parts and promoting progress of science and technology and development of the country in the end. The massive data are only

Therefore, considered from the view of work of scientific and technical information, the big data can be defined as the features of 5V: Volume, Variety, Velocity, Visualization and Value.

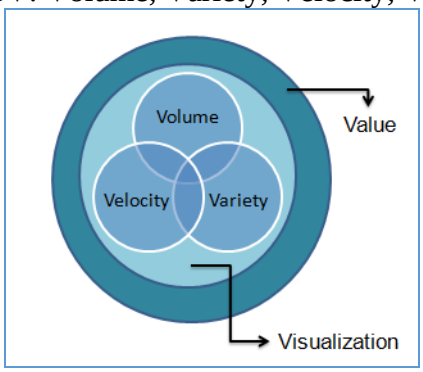

Figure 2. 5V features of big data in terms of information research

The work of scientific and technical information before the age of big data can be simply described as resources collection, application of technical approaches and providing information judgment service. In age of big data, the principle of information work doesn't change basically, but resources are replaced by 3V features big data resources (Volume, Variety and Velocity); analysis of massive big data is applied in technical approaches and visualization technology is used for data description; in the end, value of big data can be maximized via information judgment and service providing. In other words, to consider big data as a type of industry, the key of profitability of this

industry lies in improving “capacity of visualization processing” and increase the value of data via “visualization processing”.

The opportunities aroused by big data for scientific and technical information

The quiet rise of concept of “big data” is also causing the reform in society, technology, science and economy. In addition, brand new opportunities and challenges have been brought to the concept, modes, methods and technologies of information service. The information visualization in the age of big data will face the following opportunities:

1) In the age of big data, the data related to science and technology changed from digitization to datamation, which laid the foundation for science and technology searching

With development of electronic library and database of a variety of knowledge, more and more periodicals and books are being transformed to digital form and stored into the databases in PDF and other formats. However, the digitalized data are not convenient to be searched and analyzed. Firstly, you need to know which books contain the required data; secondly, you need to read the digital resources very careful to find the relevant information. This is basically the same as the way of searching for in formation in the books.

If the relevant information in the books is datamized, the words, paragraphs, information about authors, key words and other information can be identified respectively; it will be more convenient to search the information with assistant search engine of competitive information. The so-called datamation is a process to quantize an image to the analysis of tabulation. Quantization is the core of datamation and the precondition of visualization of expressed information. The potential information value of the information can not be let off until it is datamized.

On the other hand, for quantitative analysis in the previous scientific and technical information work, the most usual method is to carry out biliometric analysis of structured database; while in the age of big data, except biliometric analysis of structured data via Wanfang Data, National Science Technology Library, CNKI database, EI, Derwen Patent Database, VIP technical journals database, etc, the information analysis of unstructured data has also appeared. Big data has in fact provided the basis of massive real-time information so that the useful information will be obtained via visualization through technical methods.

As shown in the figure below, this is the visualization analysis about H7N9 public health incident happening at the beginning of 2013, the source of the data are mainly from the unstructured data of press media and official Microblogs. After the data from different sources and forms are datamized with a uniform format, analyzed with integration and visualized, the customers can see the areas with the worst conditions of H7N9 and learn the specific development conditions of H7N9 of each individual place, so that relevant determinations can be made fast and the efficiency and correctness of determination will be improved dramatically.

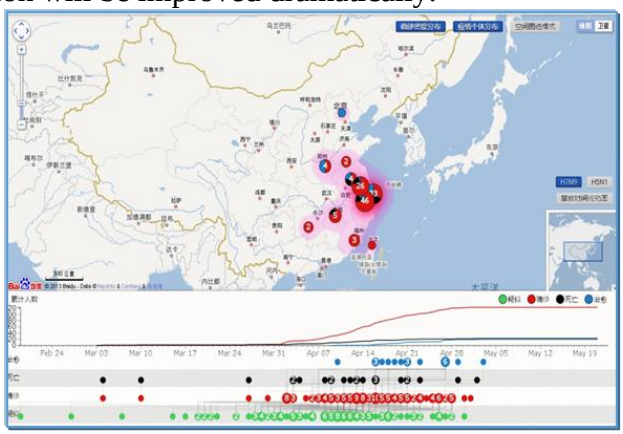

Figure 3. Visualization of H7N9 condition in China, 2013 [4]

2) The age of big data makes quantization of personal behaviors possible and provides basis for finding and learning the subject behaviors of the society

Now, smart terminals such as cell phones, computers, smart TVs, etc. are necessary in everyday life. The daily life can basically be expressed digitally. Every move of us will produce massive data. Any behavior has an omen. And on the internet are stored massive data of omens. By collecting and analyzing such data, the future human behaviors in the physical world will be predicted.

Nicholas Felton became famous in the field of "big data quantization life" for his personal annual report. The person annual reports can show the preciseness of personal data collection and his gift in design. Besides geography locations, he also tracked the persons he got along with every year, their eating places, their movies, books and other information. The figure below is one of the pages of Felton's annual personal report in 2012.

Figure 4. One of the pages of Felton's annual personal report in 2012 [5]

Since 2005, Felton has designed eight annual reports. It should be noticed that more and more private lives are exposed in his reports and the data are getting richer. As time goes by, the data are more like the diaries than the simple reports.

Although predictions with analysis of big data based on personal habits are different from person to person, the overall precision is higher than we expected. As the humans are not so dependent on internet and electronic devices before, the electronic tracks in the age of big data will make it possible to study the habits of a person, a group, or even the whole species of mankind. The possibility and reality provide some potential chances for development of scientific and technical information works and the basis of learning the behaviors of subject groups of science and technology on internet, such as experts, scientific and research institutes. The service of dynamic information of scientific and technical subjects can be provided according to the customers' requirement and basis can be provided for the development tendency of the scientific subjects; in addition, the information supply service mode will be oriented by personality and customized and will be more precise and smarter; a new mode with the performance cost ratio several times or even dozens of times higher than current information products will be created.

3) The analysis capacity of relevance of big data will make scientific and technical information closer to relevant relationship of the information objects

The intelligence transformation of information analysis is not just about the innovation of cluster and classification of computers and algorithm of relevant relationships, it's more about the huge value of the information brought by the relevant relationships. Especially in terms of identifying the external competitors, development tendency of science and technology and research direction of the scientific talents, the valuable information can be analyzed from relevant relationship.

Innovation of scientific and technical information work mode is not just the change of methods but the reform of information analysis thinking mode. "The old" pattern of scientific and technical information analysis is: confirming information fields, making plans, practicing and exploring, analyzing data and drawing conclusions. This analysis pattern is made because we know little information about the objects of analysis. In other words, we know few data about the objects to be analyzed, so we need to carry out trial analysis and study based on the few data by guessing and hypothesis and then draw conclusions. It's not about the correctness of such conclusions but the

limitation of such information. The “old” scientific and technical information analysis is based on a linear relationship between reason and result; however, when we face big data, we will find the subtle relationship between complicated events. In front of such a huge and complicated big data, a sustainable advantage in competition can be achieved only by exploring the challenges of the market and setting up flexible targets with a reformed dynamic information service system.

4) Big data has powerful prediction capacity and is good for carrying out more precise scientific and technical information service.

In the age of big data, we can analyze more data and even process all the data related to a specific event. We will not depend on the random sampling in “old” pattern. This will improve dramatically the correctness of information analysis; especially in terms of predicting the industry tendency, identifying the new competitors and industry alarming, the information analysis will be more precise; the constant tracking and monitoring to such data will find the abnormal signs, give warnings and improve timeliness of information analysis when the vector monitoring data of the key factors change abnormally.

Physicists team in Northeastern University can predict the physical location of a person at any given moment and the correctness rate is over 93%; computer scientists in Johns Hopkins University analyzed 1.6 million micro blogs about health related from May 2009 to October 2010 and found the relevance index of flue rate obtained from the data modeling and official source is 0.958 [3]. The above cases suggest that the power of data analysis is beyond natural science and business; the prediction of big data has played a role in terms of social science and public policies.

With development of society and growth of scientific and technical information demand, the scientific and technical information service is more than providing the raw information but needs to provide the valuable information after selection, information based on facts for analysis and the information for predicting future development. The precise information of high value needs the support of massive data. Only when data is accumulated to a certain degree can it show the real value.

Revolution of scientific and technical information work mode in the age of big data

In the age of information, development and application of knowledge information and degree of informationization can reflect the economy capacity of a country. China has been strengthening the construction of information system and improving economy with informationization. Scientific and technical information will play a more important role in economic construction and future development of society. With the correct and fast transmission of information, the efficiency of work and production will be improved dramatically and the determinations of science and technology and economy will be made on a feasible and stable basis. With the age of big data, the information work mode is also going through a profound change. A scientific and technical information system suitable for features of big data age is necessary for the progress of science and technology and social informationization. Only in this way can the massive big data be visualized and information analysts can carry out the further judgment of information.

With the significant features of big data age and mature information technology, the mode of information work has gone through a reform (see the figure below). In the age of big data, as the massive data can’t be judged and read manually by analysts of information, the data should be explored specifically and described visually so that the big data can be changed into small data for information workers to judge. Different from the traditional requirement of scientific and technical information work, the reformed information work needs more IT talents, who should not only learn the requirement of information and relevant knowledge but turn the data into information via IT methods and visualize and show the information to the determination makers.

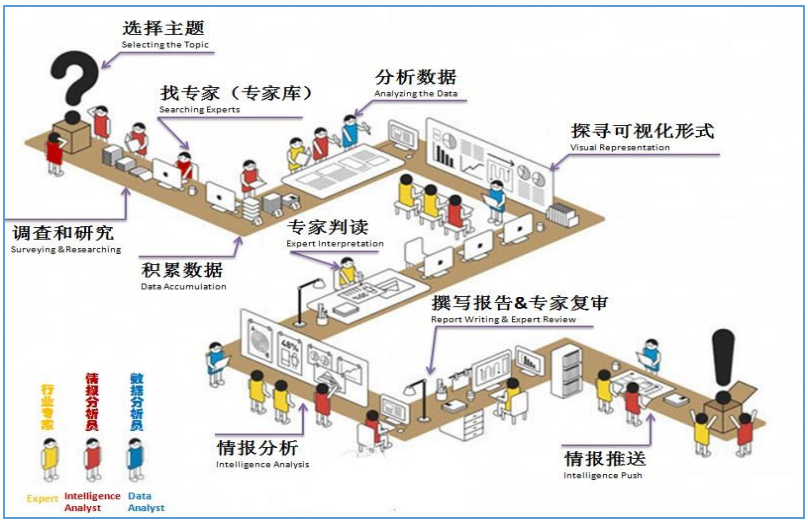

Figure 5. Reform of information work mode based on information forming process model

In the new information analysis process, the first step is to choose the topic and carry out preliminary investigation and research for the topic and interpret the topic. In addition, search for the experts in the field in expert banks and accumulate data with information search engine. These steps are carried out by information workers.

After that, data analysts and experts will join in the information work team to select and explore the collected data and research the form of visualization so that the value of the shown data will be increased. The visualization here means visualizing the description of data and is the process of turning big data into small data. After this step, experts can judge and read the visualized data description and carry out information analysis with the information analysis people so that the final information service products can be obtained. The products can be in form of report and visualized publication of gate websites.

In the process of the new information analysis, two points deserve attention:

The first one is introduction of visualization technology. As humans can't read the massive data information, visualization is necessary. Proper data exploration methods should be chosen according to requirement of information task to visualize the data description, abstract the amount of data within the reading ability of humans and analyze and judge the information. Furthermore, in form of information products pushing, the visualization technology can help the users understand the information, learn the key points correctly and make determination.

The second one is structure of information work team. The new information work team should consist of industry experts, data analysts and information analysts. It's also an information team full of features of big data.

Information analysts: determining the target of information business, planning information requirement, coordinating teamwork and analyzing and judging information;

Experts: using the knowledge, experience and judgment of experts to identifying, filtering, judging the effectiveness of the collected information; for selecting the methods of data exploration and analysis, the experts' thinking mode should be taken into consideration to analyze the data;

Data analysts: working from the view of IT, meeting the requirements of information analysts and achieving IT functions such as visualization.

In the information work team appearing in the age of big data, business language should be connected with technology language without any seam; only in this way can the application of visualization play a role in information judgment. Therefore, communication system should be established for information interpretation. In the standard of information communication system, the reformed information innovation mode suitable for the age of big data can maximize the information value. The visualization plan selected on this basis will be the most proper one.

Conclusion

Age of big data brings both opportunities and challenges to the work of information and produces the new mode of information in the new age. However, information processing and analysis is very professional work, where the data needs to be transformed into valuable information with proper data exploration plans and judged and reflected manually by experts in relevant fields. In summary, big data information resources, proper visualization plans and information judgment with artificial intelligence are the three factors supporting scientific and technical information service in the age of big data.

The topic of big data is how to bring value from the big data; the topic of scientific and technical information in the age of big data is how to use big data and produce information value correctly. In the age of big data, the information will appear at an amazing speed, so the scientific and technical information will play an important part in development of science and technology. Scientific and technical information is a serious of activities, including tracing, monitoring, analyzing and displaying the specific objects; the source is big data. For scientific and technical information, big data is a chance, a revolution and a new method to collect and interpret information and predict. Exploring the information hidden in the big data will become an important approach of industry evolution and science and technology development in China.

References

[1] Guidance about Data Center Construction and Structure, 2013, by MIIT

[2] Prediction and Analysis 2012-1016 of Big Data Technology and Service Market in China, 2011, by IDC

[3] V. M. S. &. K. Cukier, big data, 2012, p. 13.

[4] A key experiment of Beijing University machine sensing and intelligence education department http://vis.pku.edu.cn/birdflu/

[5] Nicholas Felton is one of the founders of Daytum.com and now he is a member in product design team of Facebook. http://feltron.com