Academic Journal of Business & Management

ISSN 2616-5902 Vol. 1, Issue 2: 7-12, DOI: 10.25236/AJBM.190116

Analysis of "Belt and Road" Sino-Russian trade big data ecological environment based on web crawler

Lin Xia*

Globalization and Global Issues Institute, China University of Political Science and Law, Beijing 100088, China
**Corresponding author: XIA Lin, e-mail: xialin666@126.com*

ABSTRACT. *In order to promote the "Belt and Road" Sino-Russian friendly trade, the advanced web crawler technology is used for data analysis in the context of the rapid development of Internet of Things technology. The types and applications of web crawler in big data are described; then, the web crawler technology is applied to the big data ecological environment analysis of "Polar Silk Road" of "Belt and Road" Sino-Russian trade. The steps of applying web crawler to the big data ecological environment analysis of "Polar Silk Road" are explained first; next, the analysis results are clarified. The research results have shown that the Sino-Russian trade ecological environment analysis based on web crawler technology is more perfect, which promotes the application and propagation of web crawler technology. The analysis has pointed out that being the cooperation project of Sino-Russian "Belt and Road", the "Polar Silk Road" is facing a huge test in terms of its ecological environment, in which the oil spills occurred in oil and gas transportation projects have become the most important source of pollution in the Arctic region. These ecological problems require China and Russia to strengthen the development and construction in the fields of marine biological development, navigation work, fishery market, natural energy, etc. to further promote the construction and development of the ecological environment of the "Polar Silk Road".*

KEYWORDS: *Big data; Web crawler; Technological application; Sino-Russian trade*

1. Introduction

Since the beginning of the 21st century, human society has undergone unprecedented changes. The "Polar Silk Road" is the new achievement in the development of trade between China and Russia in recent years, which has become an integral part of the overall layout of the "Belt and Road" initiative. In June 2017, the National Development and Reform Commission promulgated the *Belt and Road Construction Maritime Cooperation Plan*, which became an important achievement

ISSN 2616-5902 Vol. 1, Issue 2: 7-12, DOI: 10.25236/AJBM.190116

in the strategic development of the “Belt and Road” initiative. In the *Plan*, a total of 3 blue economic channels were clearly proposed, one of which was “The Arctic Ocean Economic Circle” connected to Europe. The promulgation of the *Plan* indicated that the “Polar Silk Road” has become an important part of the “Belt and Road’ initiative. In November of the same year, the leaders of China and Russia held a meeting. During the meeting, Chinese President Xi emphasized the importance of the “Polar Silk Road” [1].

The “Polar Silk Road” would become an important economic development link between China and Europe and affect global economic development. Due to the important commercial value of the “Polar Silk Road”, it not only helps to improve the current imbalance of regional economic development but also contributes to the foreign economic development of China, which would also promote the sustainable development of the global economy. At present, the survival and development of the “Polar Silk Road” are still facing the challenges of cost-based economic technology, geopolitical competition, and harsh natural ecological environment. In order to promote the construction of the “Polar Silk Road” and realize bilateral trade and cooperation between China and Russia, it is essential to analyze the ecological environment of the “Polar Silk Road” [2].

In the context of the rapid development of Internet of Things technology, network data is in explosive growth. The application of data requires memory-based mining and searching from a large amount of data. Therefore, search engines rise to the occasion, which not only searches data information for users but also help users find the results they need [3]. The processing of information data requires the use of web crawler technology to collect network information. The rapid excavation information technology of web crawler can better help China and Russia analyze the ecological environment of Sino-Russian trade big data. Therefore, the web crawler technology is applied to the big data analysis of the “Polar Silk Road” ecological environment, thereby comprehensively and systematically analyzing the trade ecological environment of the “Belt and Road” between China and Russia.

2. Methodology

2.1 Brief analysis of web crawler

The “web crawler” is a visual name. The crawler technology is mainly a program for automatically extracting web page information. It starts from the initial URL set and puts it all into the orderly queue to be collected. After the web crawler links to the web page, it can actively identify the HTML code of the web page. The code has the META information identifier. Under the guidance of these symbols, it can know which web pages need to be crawled and judge whether the web page links are tracked or not.

The web crawler technology is currently divided into 3 types, i.e. the focused crawler technology, the general crawler technology, and the incremental crawler technology, which are introduced below. The focus crawler technology adds a link

ISSN 2616-5902 Vol. 1, Issue 2: 7-12, DOI: 10.25236/AJBM.190116

evaluation and content evaluation module, whose main point of the crawling strategy is to evaluate the content of the page and the importance of the link. The implementation of the crawling strategy for link structure evaluation mainly takes the WEB page as a semi-structured document, which has a lot of structural information to evaluate the importance of the link. The crawling strategy based on content evaluation mainly applies the text-like calculation method, proposes the Fish Search algorithm, and uses the user input query words as the theme. With the further improvement of the algorithm, the Sharksearch algorithm can use the space vector model to calculate the page and topic relevance size [4].

The general crawler technology of the web crawler technology, also known as the whole network crawler, is a critical technology that starts with a preset single or multiple initial seed URLs to obtain the initial online URL list so that in the actual crawling process would continue to access and download the page from a URL in the queue. Then, the page content is obtained through the removal of HTML symbols by the parser; the summary is saved in the database of the WEB, and the new URL is saved in the current page until it can satisfy the conditions of system termination. The application of general crawler technology has different strategies. The breadth-first strategy and the depth-first strategy are key strategies; for example, the implementation of the depth-first strategy is to access the next-level web link in order of depth from low to high [5].

The incremental crawler technology of web crawler technology also has its distinctive features. It is an incremental update to the download page, which only crawls the newly-generated and changed web pages; therefore, the crawled pages are as new as possible. The main goal of the incremental crawling technology is to improve the quality of the page on the local page and keep the page centrally storing the latest page. The more common methods are the individual update method and the unified update method. In addition, another goal of incremental crawling technology is to rank the importance of web pages [6].

2.2 Application of web crawler in big data

The application of web crawler technology in the field of big data has distinct characteristics, which is reflected in the amount and the variety of data. The data resources are usually in all directions; thus, it is difficult to obtain data resources accurately; the application of web crawler technology could solve the problem. In terms of the application of crawler technology in the field of big data, its design is the key technology of search engine implementation and involves the crawler algorithm, which is a key element related to search engine performance [7]. The crawler algorithm plays a key role in the free search of the network system and the automatic downloading of web pages by identifying URLs.

At present, there are 3 algorithms used for crawling technology, i.e. the breadth-first algorithm, the depth-first algorithm, and the best-first algorithm. The implementation of the application strategy of the breadth-first algorithm is the time when the data information is captured. After completing the search on the current

ISSN 2616-5902 Vol. 1, Issue 2: 7-12, DOI: 10.25236/AJBM.190116

level, it enters the next level. The design of the algorithm is relatively simple and can cover as many web pages as possible. Another method is the combination of the breadth-first algorithm and the web filtering technology, which crawls the web pages through the breadth-first algorithm and then filters the unused web pages [8].

The depth-first algorithm starts with the start page, selects the URL to enter the URL in the analysis page, and repeats the processes until the line is completed then processing the next line. In the implementation of the strategy, there is great value in providing links to portals. Another example is that in the application of the best-first algorithm, the search process is to first establish a search graph G, which has an initial node set S, open=(S), with closed=() to represent the empty table; if the set open is empty, it means that the search will exit the program, n=first(open); remove(n, open); add(n, closed), n is the target node indicates that the search is successful, and can give the path from S to n; otherwise, the n-node is expanded to generate a non-n-ancestor post-set node set M={m}, and each m is added as a subsequent node of n to G; if m does not appear in the open and closed sets, then (m, open) is added; if m has a duplicate node k in the open set, then g(m) is added [9].

The best-first algorithm is to optimize the crawler technology and improve its application performance. In terms of the simulation of the best-first algorithm search, it is necessary to set A1, A2, B1, B2, B3, B4, B5 as the relevant URLs, as well as setting A2 to provide the interference factor as irrelevant web page; then, the crawler program is set to start from page A1 and capture the overall network environment coverage information; in terms of the improvement ideas, the information captured by the crawler program is compared through data analysis and filtering; if it is found that the A2 web page information and user requirements do not match, and A1 is the data information that the user needs, the crawler program will filter the information of A2 and then capture the information data of the A1 web page, which can effectively reduce the error rate of crawling information and improve the coverage and accuracy of the captured data [10].

3. Results and discussion

3.1 Application of web crawler in the big data of "Polar Silk Road" ecological environment

The process of using the web crawler technology to implement the big data analysis program for the "Polar Silk Road" ecological environment mainly includes the following steps: First, the main objectives of the ecological environment influencing factors are locked. The target data that users want to obtain are determined based on the targets, and the main web pages that these data are distributed are identified. Second, the target web page is analyzed. In order to capture relevant data more efficiently, the structural analysis of the target web page is mainly the path and logic analysis of its data access; then, the data is captured. The obtained data results are saved through the execution of commands by the selected software. Finally, data analysis is performed. The data obtained through the

ISSN 2616-5902 Vol. 1, Issue 2: 7-12, DOI: 10.25236/AJBM.190116

above program is cleaned to obtain the target data, which can be further analyzed, expanded, and confirmed through the combination of technologies such as SQL query, Excel analysis, and data visualization, etc.

The principle of applying web crawler technology to implement data analysis is to obtain more sufficient and high-quality target data that are more reliable from the outside of the locked target, and the acquired new data can be compared with the existing internal data to help analysts expand the scope and find more relevant data results, which makes the analysis more reliable.

3.2 Analysis of big data of "Polar Silk Road" ecological environment

The previous section has described the application steps of the web crawler technology in the "Polar Silk Road" of "Belt and Road" Sino-Russian trade big data ecological environment. The analysis results obtained from the technology application are explained below.

The natural ecological environment of the Arctic region on the "Polar Silk Road" is very fragile, and the navigation conditions are poor. At the same time, there are certain differences in geographical advantages and characteristics near the Arctic Ocean, which brings more challenges and pressures to the navigation activities of various countries. In addition, the important role of the Arctic waterway is to provide countries with energy output and commercial expansion space; thus, the relevant ships must operate in accordance with the provisions of the international maritime department. China, as a flag state but not an Arctic country, will face serious constraints on the development and application of the northern marine channels. Besides, Russia also faces the same problem. Although Russia is an Arctic country that can act for a long-term in the Arctic Ocean and has rich experiences of thecomplex ice conditions in the sea, the icebreakers and transport vessels used in the Russian northern marine channels are too long and basically do not meet the requirements of environmental protection; therefore, technology updates, ship updates, and domestic legal amendments are also challenges for Russia.

In view of the ecological problems existing in the above analysis, China and Russia need to make efforts in the following areas, including promoting the development of marine development in the Arctic region, navigation work, fishery market, etc.; in addition, China needs to make important contributions in the fields of navigation channel management, ship construction, marine technology, and personnel services in the Arctic region; besides, transportation service stations should be set up in the Arctic Ocean to provide navigation services for vessels.

China is the largest trading country worldwide and urgently needs to develop transportation routes in the Arctic Ocean region. China and Russia should cooperate in the construction of transportation infrastructure in the Arctic Ocean region, and improve the transportation conditions in the Arctic Ocean region to facilitate marine transportation in China and other countries.

ISSN 2616-5902 Vol. 1, Issue 2: 7-12, DOI: 10.25236/AJBM.190116

4. Conclusion

In summary, under the current development background of the information age, the amount of data is in explosive growth, and the requirements of information application of the public have been increasing. Therefore, corresponding technologies are required to provide convenience for people, while the application of web crawler technology can meet the needs of the public. The web crawler technology is applied to the analysis of the Sino-Russian trade ecological environment of the “Belt and Road” with the construction of “Polar Silk Road” as the example; the analysis results have shown that as an important part of the overall “Belt and Road” initiative of China and Russia and the new achievement in the development of trade between the two countries, the “Polar Silk Road” is facing enormous challenges in terms of ecological environment, which requires both sides to make great efforts in the marine biological development, navigation work, fishery market, and natural energy, etc. to jointly promote the Sino-Russian friendly trade exchanges.

References

[1] Khaliduzzaman A, Fujitani S, Kashimori A, et al. A non-invasive diagnosis technique of chick embryonic cardiac arrhythmia using near infrared light. Computers and Electronics in Agriculture, 2019, 158, pp. 326-334.
[2] Huang Y. Understanding China's Belt & Road initiative: motivation, framework and assessment[J]. China Economic Review, 2016, 40, pp. 314-321.
[3] Rowatt K, Burns R E, Frasca Jr S, et al. A combination Prussian blue–hematoxylin and eosin staining technique for identification of iron and other histological features. Journal of Histotechnology, 2018, 41(1), pp. 29-34.
[4] Yang H, Zhao X, Han B, et al. Spatiotemporal patterns of Amur leopards in northeast China: Influence of tigers, prey, and humans. Mammalian Biology, 2018, 92, pp. 120-128.
[5] Ma L, Sun G, Qu Y, et al. Applying systematic conservation planning to constitute a protection strategy for broad-leaved Korean pine forests in Changbai Mountains, China. Journal of Mountain Science, 2016, 13(3), pp. 493-507.
[6] Liu W, Dunford M. Inclusive globalization: Unpacking China's belt and road initiative. Area Development and Policy, 2016, 1(3), pp. 323-340.
[7] Huang Y. Understanding China's Belt & Road initiative: motivation, framework and assessment. China Economic Review, 2016, 40, pp. 314-321.
[8] Wen X, Ma H L, Choi T M, et al. Impacts of the Belt and Road Initiative on the China-Europe trading route selections. Transportation Research Part E: Logistics and Transportation Review, 2019, 122, pp. 581-604.
[9] Unpacking China's belt and road initiative. Area Development and Policy, 2016, 1(3), pp. 323-360.
[10] Chan M H T. The Belt and Road Initiative–the New Silk Road: a research agenda. Journal of Contemporary East Asia Studies, 2018, 7(2), pp. 104-123.