Northeast Asia as Seen Through Data Analytics

Ilhong Ko

In recent years, 'Big Data' and 'data analytics' have become the buzzwords of the research community. Some disciplines have been active in employing new digital technologies, while other disciplines have been less inclined to do so. In the case of Asian Studies, researchers specializing in East Asia have played a pivotal role in bringing about a 'Digital Turn' and are now actively engaged in several Big Data-related initiatives. In this issue of News from Northeast Asia, we examine how researchers based at Seoul National University Asia Center (SNUAC), faced with the need to develop flexible and open-ended approaches to fieldwork amidst the COVID-19 pandemic, have utilized text analysis as a useful research tool to gain new insights into the region of Northeast Asia.

Big Data analytics on Asia-related news articles can provide information on the key topics of interest concerning Asia, on how the topics of interest differ from country to country, and on how they change over time. This is demonstrated by Myungmoo Lee of Seoul National University Asia Center and Dohoon Kim of Ars Praxia in "Big Data Analytics of Northeast Asia's Top 10 News Topics." By identifying the common concerns shared by the countries of Northeast Asia, the results of the study may provide the basis for enhanced cooperation in the region.

Text analytics can also be usefully applied to smaller data sets that do not fall within the category of 'Big Data,' one such example being the contents of UNESCO's World Heritage List website. "Examining Asia's Cultural Heritage on the UNESCO World Heritage List through Data" presents the results of text analytics that Minjae Zoh, Ilhong Ko, and Junyoung Park of Seoul National University Asia Center undertook on the 'Outstanding Universal Value (OUV)' content of 187 World Cultural Heritage Sites located in Asia. The results provide insights into both the distinctive and the universal aspects of the World Cultural Heritage Sites of Northeast Asia.

Ilhong Ko, HK Research Professor, Seoul National University Asia Center.
Email: mahari95@snu.ac.kr

SNUAC
Seoul National University Asia Center

The Seoul National University Asia Center (SNUAC) is a research and international exchange institute based in Seoul, South Korea. The SNUAC's most distinctive feature is its cooperative approach in fostering research projects and international exchange program through close interactions between regional and thematic research programs about Asia and the world. To pursue its mission to become a hub of Asian Studies, SNUAC research teams are divided by different regions and themes. Research centers and programs are closely integrated, providing a solid foundation for deeper analysis of Asian society.

Big Data Analytics of Northeast Asia's Top 10 News Topics

Myungmoo Lee and Dohoon Kim

Text analytics are widely used to extract information and patterns from text. By applying a combination of such text analytics techniques to newspaper articles, priorities and patterns in the news can be identified. In an attempt to establish the Asia-related issues that concerned the countries of Northeast Asia the most in recent years, newspaper articles were analyzed by researchers based at Seoul National University Asia Center, working in association with staff from of Ars Praxia, a company specializing in big data analytics and digital contents creation. The analyzed data consisted of Asia-related news articles published in the English-language newspapers of South Korea, China, Japan, and the United States between January 1, 2020 and September 30, 2022.

The original data set comprised a total of 5,502,266 articles from 824 news outlets. In order to mitigate the bias arising from the fact that 4,796,149 of these articles came from American news outlets, additional sampling was undertaken. The data set that was analyzed consisted of the following number of articles: 46,036 (South Korea); 76,171 (China); 44,298 (Japan); 94,521 (USA). Topic analysis was undertaken on these articles using topic modeling, trend analysis, and semantic analysis techniques.

Using topic modeling, which is a machine learning technique that that analyzes unstructured text data in order to recognize latent topics, the top ten Asia-related news topics were identified and then their relative importance was established for each country. All four countries showed a high interest in economic crisis and energy related crisis. In addition, the North Korean nuclear crisis was commonly regarded as a topic of relative high interest in South Korea and Japan. On the other hand, cross-strait issues and security crises (which included issues such as tension in the Taiwan Strait and fishery disputes) were only a high-interest topic for Chinese news outlets.

Then how did interest in these ten topics change over time? In order to explore this question, trend analysis was carried out. The article frequency of each topic was calculated per quarter and changes in the frequency were traced over time. As illustrated in Figure 1, the number of articles on the COVID-19 pandemic was overwhelmingly high, with an average of 214,000 articles per quarter. However, in terms of relative acceleration, this topic had the lowest negative value (-3751.17), suggesting that interest in this topic will soon disappear. The economic crises of major Asian countries, on the other hand, is a topic characterized by both high frequency (130030.55) and acceleration (2735.57) in terms of average articles per quarter. This indicates that the countries of Northeast Asia and the United States expect this issue to become a constant threat in the near future. The Russo-Ukrainian war, believed to play a pivotal role in determining the direction of the international order in 2023, illustrated the third highest average frequency (50954.27) and the highest acceleration (3001.56).

Semantic network analysis was used to identify the keywords of interest, as well as the relationships between these keywords, for each the four countries. The semantic network of the articles from the news outlets of all four countries reveals that 'engine' (of economic growth), 'COVID strategy,' 'energy supply,' 'technology,' 'cruise missile,' and 'stock price' were the most frequently occurring keywords [Fig. 2 see overleaf].

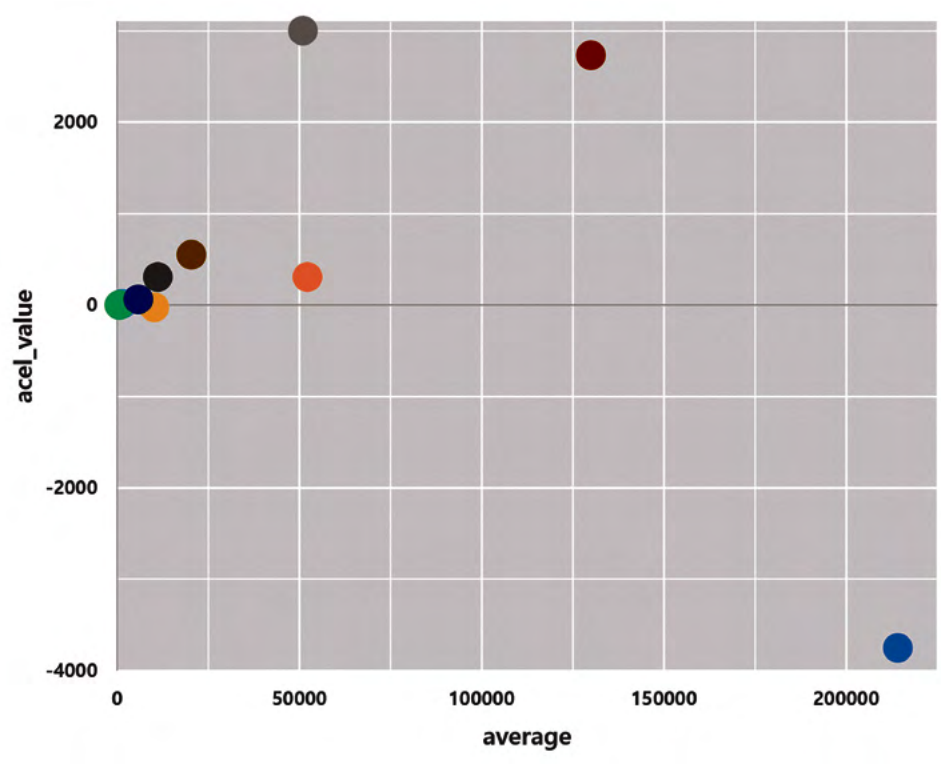


Fig. 1: Article frequency (X-axis) and acceleration (Y-axis) for the top ten topics (January 1, 2020 to Sept. 30, 2022). (Figure by the authors, 2023)

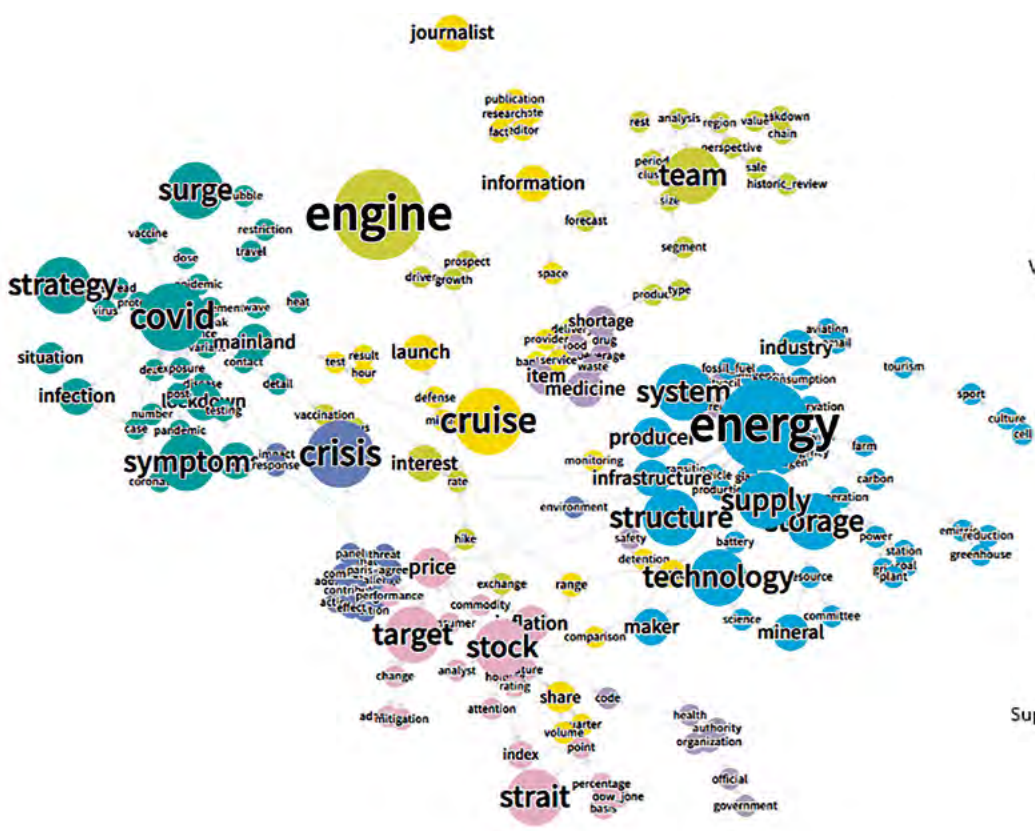
Continued from previous page

Using the nodes (keywords) with the highest betweenness centrality values, the shared or unique key interests of each of the four countries were then identified and visualized, as presented in Figure 3. It can be observed that 'energy,' located in the center, was the key interest shared by all four countries. Common key interests are expressed as nodes placed between the respective countries, and unique interests are expressed as dots located along the periphery of each country node. Inflation (represented by 'price') was a common concern to the three Northeast Asian countries, and South Korea and Japan also shared concerns about the missiles launched by North Korea. It is of interest to note that, for China, 'health' was identified to be a key interest, alongside 'development,' 'security,' and 'trade.' This suggests that the government's handling of the pandemic at home has come to be regarded as a key issue affecting government stability.

These results of Big Data analytics clearly show the complex security risks that have emerged due to the geopolitical shifts of the post-COVID-19 era. With the possibility of escalating military conflicts, such as those between Russia and Ukraine, the concept of 'security' has expanded and evolved to include responses to climate and environmental crises, public health crises, the securing of value chains, protection of technological competitiveness, and well-controlled trade and economy. It will be interesting to trace the predictive nature of these results of big data analytics in order judge the viability of Big Data analytics as an alternative research tool for gaining insights into the region.

Myungmoo Lee
Seoul National University Asia Center.
Email: leemm@snu.ac.kr

Dohoon Kim
Ars Praxia.
Email: leo_kim@arspraxia.com



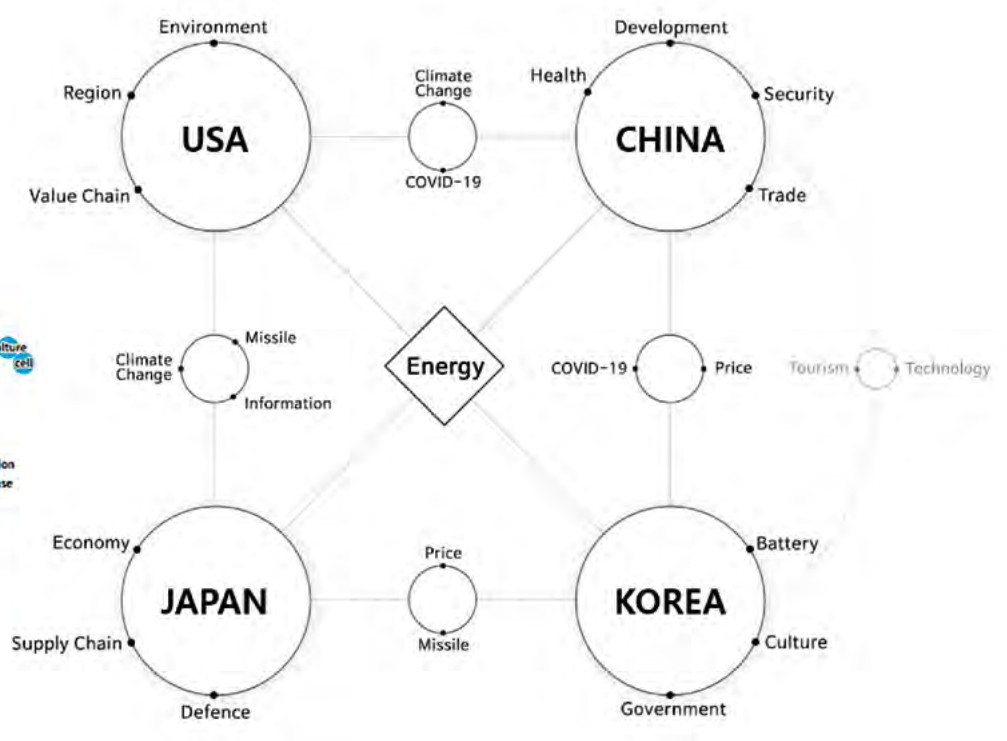


Fig. 2 (below left): Semantic network of Asia-related news articles from the English news outlets of South Korea, China, Japan and the USA (January 1, 2020 to Sept. 30, 2022). (Figure by the authors, 2023)

Fig. 3 (below right): Shared or unique key interests as seen through Asia-related news articles from the English news outlets of South Korea, China, Japan and the USA (January 1, 2020 to Sept. 30, 2022). (Figure by the authors, 2023)

Examining Northeast Asia's Cultural Heritage on the UNESCO World Heritage List through Data

Ilhong Ko, Minjae Zoh, and Junyoung Park

Fig. 1: Word cloud illustrating the most frequently occurring words in the official texts related to all of Asia's Cultural World Heritage Sites (n=187). (Figure by the authors, 2023)

The concept of 'World Heritage' was conceived and developed in the West by the UNESCO Organization in 1972, and since the establishment of the official World Heritage Convention, almost all Asian countries have actively designated their national heritage as 'World Heritage.' In principle, 'universalism' is at the core of the World Heritage Convention's concept of 'Outstanding Universal Value (OUV).' However, because this framework was shaped and cemented in the West, this means that Western standards were and continue to be applied to other parts of the world, including Asia. With this in mind, data analytics can be used to question whether a pattern can be found to decipher a particular 'Asian' characteristic or strategy in the case of the sites of Northeast Asia.

According to the World Heritage Convention, the region of 'Asia' is categorized under 'Asia and the Pacific.' As of 2021, a total of 187 cultural heritage sites were registered within the 'Asia and the Pacific' region. The findings from quantitative data analysis reveal that the inscription of sites located in Asia began in 1979, a year after the first set of World Heritage Sites became designated in 1978. Iran (with three Cultural Heritage Sites) and Nepal (with one Cultural Heritage Site) were the first countries to designate. With the turn of the 1980s, South Asia was responsible for pushing up the inscription rate in Asia and then, in the 1990s, Northeast Asia began to govern the fluctuation patterns of the region. From that point forward, the fluctuation patterns for site inscriptions in Northeast Asia are mirrored by the fluctuation patterns for the whole of Asia.