Quantitative Content Analysis of Japan's Cybersecurity Strategies since 2013

Piyush Ghasiya,
Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka

Prof. Sachio Hirokawa,
Research Institute of Information Technology, Kyushu University, Fukuoka

Prof. Koji Okamura,
Research Institute of Information Technology, Kyushu University, Fukuoka

Abstract

Japan has published three “Cybersecurity Strategy”[1] since 2013. [1] [2] [3] This research would include Quantitative Content Analysis (QCA) of all three cybersecurity strategies since 2013. “Lexicometrics” approach of Text Mining will be used in QCA and for that KH Coder will be used. QCA will help in locating the similarity and differences in the Cybersecurity Strategies over the years.

Keywords

Text Mining, Lexicometrics, QCA, Cybersecurity, Japan

I. Introduction

If seriousness towards an area or field or issue is gauged by the publishing of the public policy document, then in the field of cybersecurity, Japan is the most serious country in the world. Since 2013, Japan has published three “Cybersecurity Strategy.” These strategies give an overview of Japan’s approach, toward cybersecurity. These documents are critical in understanding Japan’s objectives, concerns, and plans to overcome the shortcomings related to cybersecurity. This research paper deals with the Quantitative Content Analysis (QCA) of all three-cybersecurity strategy (2013, 2015, 2018) documents. KH Coder [4] is used for the QCA.

II. Methodology

Broadly speaking there can be two main text Mining (TM) applications namely “Lexicometrics” and “Machine Learning” (ML). [5] Both approaches can be used for QCA. This paper will use the Lexicometrics approach for QCA. Lexicometrics approaches aim to identify major semantic structures inductively in digital text collections. It includes four methods: Frequency analysis, Key term extraction, Co-occurrence analysis (or Collocation analysis), and Dimension reduction. Dimension reduction is useful when there are different variables in the text for example chapters, but since the cybersecurity strategies do not have different variables, it cannot be applicable here.

III. QCA of Japan’s Cybersecurity Strategies Since 2013

Japanese government published its first Cybersecurity Strategy in 2013. Before that, since 2006 Japan has continuously published “Information Technology Strategy.” The last “Information Technology Strategy” came out in 2011. Below is the passage from Japan’s first Cybersecurity Strategy which explains the reason of the Japanese government for shifting from “Information Technology” to “Cybersecurity.”

“Thus, this strategy has been named the "Cybersecurity Strategy" in order to make clear the necessity to widely promote measures related to cyberspace and approach of these measures as distinguished from efforts for assuring "information security" up until now.” [6]

IV. Frequency Analysis

As the name suggests, Frequency analysis is a method where the frequency of each term (can also be called concepts or words) is counted (in this paper by using KH Coder) and try to find how different the usage of terms are in each document. This method gives an overview of the documents. Below are some inferences from Frequency Analysis.

1. In 2013, cybersecurity was not in the top 15 list. Rather the word “information” tops the list, and if we add the word security and Security, we find that it became the second most frequent word. This shows that “Information Security” is still the focus of the “Cybersecurity Strategy” of 2013. This also points out the ongoing transition from “Information Security” to “Cybersecurity.”
2. Further, in 2013, the other main concepts that are in the top 15 include international and critical infrastructure. The appearance of these concepts shows the focus on critical infrastructure and looking for cooperation internationally.
3. In 2015, cybersecurity reached the third spot showed the completion of the transition which started in 2013. Presence of the word business shows that the Japanese government is focusing on businesses or private sector for achieving balance in managing the cybersecurity. The word International is again present in the list.
4. In 2018, there is not a major concept present in the top 15 list.

V. Co-occurrence analysis

Co-occurrence analysis is based on the hypothesis that co-occurring entities (here words) are functionally linked. [7] When visualize it shows a network of words which occurred together in a sentence or with a left/right neighbor.

With KH Coder, visualization of Co-occurrence network includes subgraph and centrality. In this paper, a centrality matrix is used. Centrality is a concept that shows which concept is central to the network. KH Coder gives the option of three centrality matrix namely Degree, Betweenness, and Eigenvector to measured the centrality. In this paper, the researcher has used the Betweenness

[1] This paper is using the English provisional translation of Cybersecurity Strategy provided by government of Japan for all three documents.

matrix for visualization. Betweenness matrix shows which nodes (words) are significant as a bridge.
Figures 1, 2, and 3 show Co-occurrence Network[2] of Cybersecurity Strategy of 2013, 2015 and 2018 respectively.

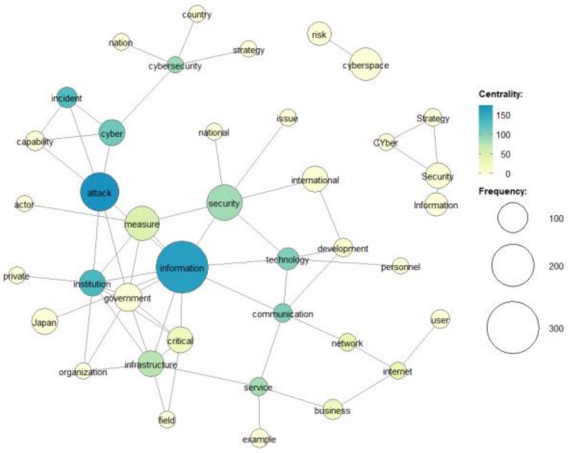

Figure 1: Co-occurrence Network of Cybersecurity Strategy of Japan, 2013.

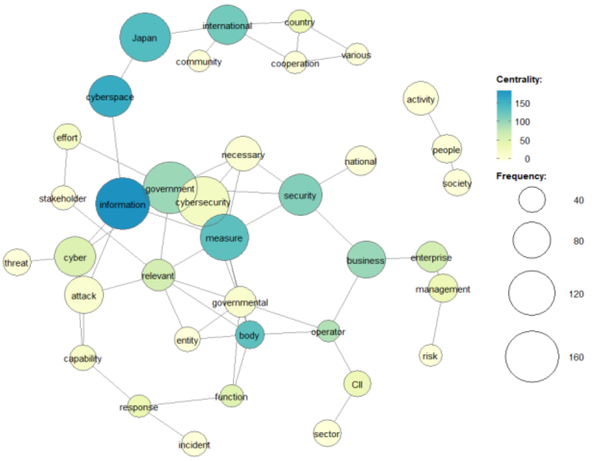

Figure 2: Co-occurrence Network of Cybersecurity Strategy of Japan, 2015.

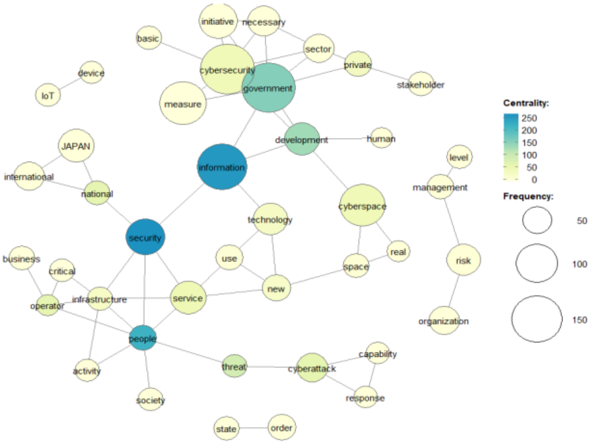

Figure 3: Co-occurrence Network of Cybersecurity Strategy of Japan, 2018.

VI. Consideration

The significance of nodes decreases from dark blue to lighter colors. Though there are few nodes in dark blue color all three cybersecurity strategies "information" is common. This demonstrates the centrality (significance) of the word "information" in Japan's cybersecurity strategies. Other than "information," "attack" and "institutions" were the critical words in 2013. Similarly, "cyberspace" and "measure" were the important words in 2015, and in 2018 it was "security," and "people." We can infer that attention has been given to these specific words (concepts).

VII. Conclusion

Frequency Analysis and Co-occurrence network analysis shows that the word 'information' is significant for Japanese cybersecurity strategies. This can be seen as a link between cybersecurity and its (sort of) predecessor "information security." These inferences can be further analyzed to give a more clear understanding of Japan's cybersecurity strategies. One such point is to find out the time taken for the transition from 'information security' to 'cybersecurity.' This point will be dealt with in the future work.

VIII. Acknowledgment

This research was supported by JSPS KAKENHI Grant Number JP16K00480.

IX. References

[1] G. ISPC, "NICS," 10 June 2013. [Online]. Available: https://www.nisc.go.jp/eng/pdf/cybersecuritystrategy-en.pdf. [Accessed 08 December 2018].

[2] GOJ, "NISC," 04 September 2015. [Online]. Available: https://www.nisc.go.jp/eng/pdf/cs-strategy-en.pdf. [Accessed 13 December 2018].

[3] GOJ, "NISC," 27 July 2018. [Online]. Available: https://www.nisc.go.jp/eng/pdf/cs-senryaku2018-en.pdf. [Accessed 13 December 2018].

[4] K. Higuchi, "KH Coder Index Page," GitHub, 2018. [Online]. Available: http://khcoder.net/en/. [Accessed 08 December 2018].

[5] G. Wiedemann, Text Mining for Qualitative Data Analysis in Social Sciences, Wiesbaden: Springer VS, 2016.

[6] GOJ. ISPC, "NICS," 10 June 2013. [Online]. Available: https://www.nisc.go.jp/eng/pdf/cybersecuritystrategy-en.pdf. [Accessed 08 December 2018].

[7] M. F. Müller H, "Identification and Analysis of Co-Occurrence Networks with NetCutter," 10 September 2008. [Online]. Available: https://doi.org/10.1371/journal.pone.0003178. [Accessed 08 December 2018].

[2] Parameters taken for Co-occurrence Network are: Term Frequency 30 or more; POS: Noun, Proper Noun, and Adjective; Centrality: Betweenness.