UTILIZING CHAT GPT FOR TEXT MINING USING DESCRIPTIVE INFORMATION OF TUNNEL FACE EVALUATION

Faculty of Environmental and Urban Engineering, Kansai University Regular Member ○Takafumi Kitaoka
Faculty of Environmental and Urban Engineering, Kansai University Regular Member Jumpei Kagamido
TAISEI Corporation Regular Member Shuntarou Miyanaga
TAISEI Corporation Regular Member Kazuo Sakai
Faculty of Societal Safety Sciences, Kansai University Regular Member Tomofumi Koyama

1. Introduction

The enormous number of papers, reports, and other publications in the construction field makes it difficult to quickly obtain a broad view of all the contents. Technologies utilizing artificial intelligence are currently in the validation phase in many fields aiming at improving productivity. The data used for analysis by artificial intelligence includes not only images and numerical values but also sound, language, and other information. In the case of language, artificial intelligence is used for natural language processing, including translation, text summarization, and automatic text generation. For a mountain tunnel, a face observation book is created in which entries are organized. Important information on the state of the face is described in the entry fields of a face observation record (For example, in case of any abnormalities occurring during operation), but it is difficult for off-site engineers to review all these records. This problem could be overcome by effectively using images created by text mining to comprehensively grasp the geological conditions, but there are few studies available for this purpose. Especially in a mountain tunnel, the natural ground cannot be visually checked upon completion, and therefore it is important to properly record information while the work is executed. Amidst the acceleration of globalization, it is expected that visualizing important information from a large amount of textual data through text mining in other languages will make it easier to overcome language barriers and bring further added value to the construction industry. The authors reported their success in creating a correlation diagram, through text mining, that can be used to grasp an overview of the natural ground conditions visually and instantaneously[1]. However, there have been few studies on text mining in this field. Therefore, in this study, we conducted a comparative analysis of results created in Japanese and English, using text mining as an example, and added some discussion. For English, we used Chat GPT for translation.

2. Research methods

2. 1 Overview of Text Mining

Text mining, one of the most notable technologies in natural language processing, is a tool that quantitatively analyzes language. Taking advantage of images based on text mining, they can be used for maintenance. Figure 1 shows the procedure for text mining. Japanese language text is written without spaces, and thus, sentences must first be broken down into words by morphological analysis. This is followed by searching, tabulation, analysis, and visualization. Words are classified into parts of speech by morphological analysis, and word-to-word relationships are analyzed and visualized, which is called a co-occurrence network. Various methods, such as correspondence analysis, are used to perform analyses in combination with properties. The present research used the KH Coder 3 software developed by Higuchi[2].

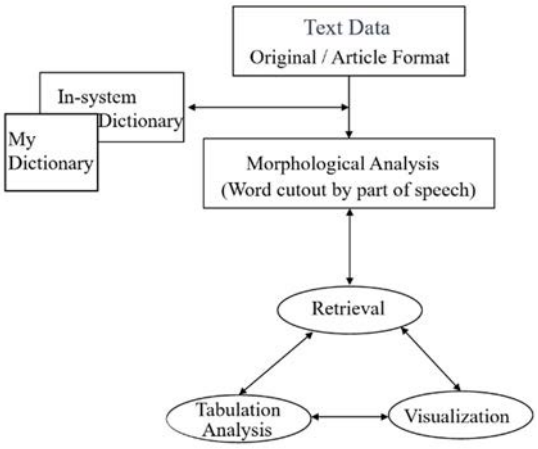


Fig.1 Text Mining Procedure

Keywords: Text Mining, Chat GPT, AI, Mountain tunnel, Tunnel Face
Contact address: 3-3-35 Yamate-cho, Suita-shi, Osaka, 564-8680 Japan, Tel: +81-6-6368-1922

2．2 Translation by Chat GPT

Figure 2 shows an example of a translated article from tunnel face. The correspondence is in chat format, and by specifying that you want to translate a sentence "using English words in the construction industry", it can correspond to technical terms, increasing the reliability of the translation. However, please be aware that sometimes the word "andesite" may suddenly be translated into Japanese, or may be translated as "basalt".

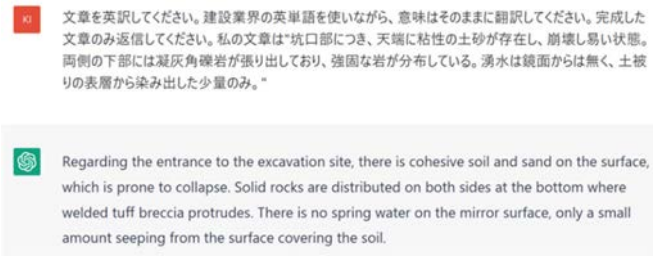


Fig.2 An example of ChatGPT translation

2．3 A face observation book

The data used for analysis in the present research consisted of entries written in a face observation book. The present research performed an analysis from the data of a face observation book created during the construction of the Atoda Tunnel (east section) on National Route 212 in Oita Prefecture. In the entries written in the face observation book, 47 faces were organized and analyzed for each rock type (andesite and tuff breccia).

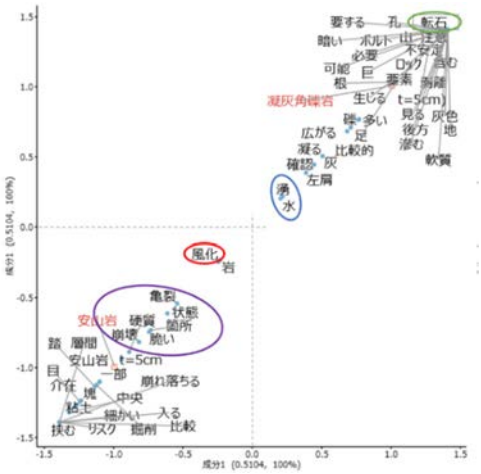
3 Correspondence analysis

3．1 Outline

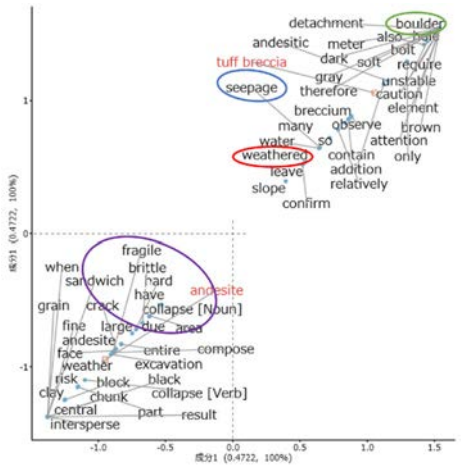
Correspondence analysis is used to find the correspondence between entries and external variables. Here, andesite and tuff breccia were set as external variables to classify the characteristics of the two rock types.

3．2 Results and Discussion

Figure 3 shows the results of the correspondence analysis. As shown in the figure, the expressions “硬質” (hard) and “崩壊” (collapse) are scattered in the andesite section, whereas in the tuff breccia section, “軽石” (boulder) and “不安定” (unstable) are used to express the geological conditions. (refer to the purple frame and green frame). In addition, it can be understood that the term "weathering" has both properties, as shown in the central part (refer to the red frame). The same applies to "spring water" (refer to the blue frame). The effectiveness of text mining is that it is easy to see at a glance that andesite, which was evaluated as hard in the geological survey report, actually had collapsed and weathered sections when excavated, and that the tuff breccia was generally consistent with prior predictions, such as multiple fractures.



a) Japanese



b) English

Fig.3 Results of Correspondence Analysis

4．Summary

The present research conducted a quantitative analysis on the entries written in a face observation book through text mining and then presented a proposal for effectively using the text information. Note that A face observation book is a record that targets face, so it is not compiled by section with characteristics like geological survey reports, etc. It is expected to be even more useful if text mining can add information on the occurrence sections or ranges.

References

1) Higuchi, K. (2016). A Two-Step Approach to Quantitative Content Analysis: KH Coder Tutorial Using Anne of Green Gables (Part I). Ritsumeikan Social Science Review, 52 (3), 77-91.
2) Takafumi, K., Chin, K., Shuntarou, M., Kazuo S. and Tomofumi K. (2023). Attempt at Quantification by Text Mining using Descriptive Information of Tunnel Face Evaluation, 49th Symposium on Rock Mechanics, 145-149.