

Visualization for comparative analysis of evaluation of licensed nursery schools by educational experts

Rika Tarumi
Ochanomizu University
Tokyo, Japan
tarumi.rika@is.ocha.ac.jp

Asahi Hentona
CyberAgent
Tokyo, Japan
hentona_asahi@cyberagent.co.jp

Karsten Klein
University of Konstanz
Konstanz, Germany
karsten.klein@uni-konstanz.de

Takayuki Itoh
Ochanomizu University
Tokyo, Japan
itot@is.ocha.ac.jp

Abstract—In recent years, licensed nursery schools in Japan have been evaluated from various perspectives. The Ministry of Health, Labour and Welfare (MHLW) encourages local governments and operating organizations to undergo third-party evaluations to improve staff morale and gain parental trust. However, the results of these evaluations are usually presented as plain text for each item. This makes it difficult to compare and interpret the data intuitively. To address this issue, we developed a visualization system to support efficient analysis and understanding of third-party evaluation data. The system is designed for educational specialists from organizations and local governments, rather than childcare professionals. We collected evaluation data for licensed nursery schools in Bunkyo-ku, Tokyo, and applied clustering to group similar schools. This system maps the results using color coding to enhance visual clarity. We also implemented co-occurrence network analysis to visualize key expressions and reveal similarities and differences among nursery schools. This system improves the readability of evaluation content and enables intuitive comparative analysis across multiple nursery schools.

Index Terms—Visualization, Machine learning, Network, Clustering

I. INTRODUCTION

In recent years, the diversity of early childhood education in Japan has expanded with changes in society, and the evaluation criteria for early childhood education have diversified accordingly. In Japan, the Social Welfare Council (SWC) has established guidelines for third-party evaluation standards. Evaluations of licensed nursery schools are conducted based on these guidelines using 34 specific evaluation criteria [1]. According to the Ministry of Health, Labour and Welfare, local governments and organizations that operate licensed nursery schools are expected to receive multifaceted evaluations by third parties (third-party evaluations) to improve staff morale and gain the trust of parents by publicizing evaluation results [2].

However, these data of third-party evaluation results present challenges for educational specialists, including municipalities and organizations that operate licensed nursery schools, in extracting and understanding the critical features of schools. The evaluation results, which are made available on online platforms, have low readability. These data are organized as separate pages for each nursery school and presented in text-based formats based on multiple evaluation scales [3]. Furthermore, it is not easy to compare their own licensed

nursery schools with those of other facilities, such as how they are evaluated compared to other schools, what points are highly evaluated, and conversely, what points are considered to be problems.

To address issues of this nature, various information visualization techniques have been proposed for unstructured text data obtained from websites to facilitate visual understanding. Hirsch et al. [4] developed txt2vz, a tool that uses co-occurrence networks to automatically generate visual summaries of textual data obtained from documents, websites, and SNS (Social Networking Service) feeds. Maarten [5] proposed “BERTopic,” which enables clustering of text data in a context-aware document embedding space, and can generate point-by-point topic representations in response to time-series data posted on SNSs.

Here, this study utilizes a limited and small-scale dataset of third-party evaluations of licensed nursery schools, specifically focusing on those located in Bunkyo Ward, Tokyo, Japan. Given the constraints of the data, we developed a visualization system by referring to existing studies while shifting the analytical focus from individual childcare workers to educational professionals, such as the organizations and municipalities that manage nursery schools.

The system is designed to support the practical tasks of educational professionals by enabling comparative analysis of evaluation results across multiple licensed nursery schools. The ultimate goal is to assist decision-makers in reviewing and refining the management policies of the nursery schools they operate. Furthermore, the system incorporates clustering methods and co-occurrence network visualizations to handle the vast amount of unstructured textual data available on the web. These techniques aim to emphasize contextual understanding and enhance interpretability and intuitive comprehension of the information presented.

II. RELATED WORK

Existing studies on the analysis and visualization of childcare-related data are limited, and there are few examples of the application of visualization methods based on text data of nursery school in particular. Therefore, this study aims to bring a new perspective to the visualization of evaluations of nursery schools by applying the text data analysis approach

adopted in various fields as well as clustering methods and network visualization.

A. Clustering Methods

There have been many studies on visualization methods for context-aware clustering. Eklund et al. [6] systematically evaluated the impact of configurations (vectorization, dimensionality reduction, clustering) and their parameter combinations, on clustering performance. As a result, BERT (Bidirectional Encoder Representations from Transformers), UMAP (Uniform Manifold Approximation and Projection) ($n_neighbors \approx 20$) and HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) were recommended, and 15-25 dimensions were preferred. In our study, we also pointed out that PCA (Principal Component Analysis) as a dimension reduction method and k-means clustering as a clustering method are also effective when interpretability is important.

Furthermore, there have been studies on review analysis using clustering methods in the field of education. By combining topic modelling using “BERTopic” and polarity analysis using “LSTM (Long Short Term Memory)”, Zhang et al. [7] developed a framework for evaluating the quality of digital educational resources based on online reviews by learners.

In this study, we conducted comparative experiments on document clustering configurations and an identification of a suitable configuration for nursery evaluation data.

B. Network Visualization

Analysis and visualization of textual data using co-occurrence networks has been widely used in various fields.

Jayaraman et al. [8] extracted information on product features and review impressions (positive or negative) from camera product reviews and visualized which features were most influential using co-occurrence networks. Similarly, our study performs polarity analysis to achieve a visualization that allows the user to understand at a glance whether a node is positive or negative. Kim et al. [9] also compared the teaching contents of children with severe and multiple disabilities in special-needs schools using co-occurrence networks, and revealed the similarities and differences in the teaching contents. Similarly, our study compares evaluation data from different licensed nursery schools, enabling a visualization of the similarities and differences between the schools.

Existing studies introduced in this section show the effectiveness of extracting features from textual data and visualizing their relationships. Our study applies such a methodology to third-party evaluation data of licensed nursery schools and aims to construct a system that enables comparative analysis of licensed nursery school evaluations. In particular, by visually presenting the similarities and differences while comparing multiple nursery schools, our goal is to clarify the evaluation characteristics of each nursery school and to support decision-making by education specialists.

Furthermore, comparative visualization has been studied in various domains such as email communication analysis [10] and output evaluation of large language models (LLMs) [11].

These works highlight the importance of designing visualizations that facilitate side-by-side comparison of complex text-based data. However, most of these approaches are tailored to specific domains and cannot be directly applied to the context of nursery school evaluations. In this regard, our study is aligned with general comparative visualization guidelines proposed by Gleicher et al. [12], especially in the use of consistent layout and color schemes to highlight similarities and differences across network representations. Our proposed system is, to the best of our knowledge, one of the first to apply these principles to comparative analysis of third-party evaluation data in early childhood education.

III. PROPOSED VISUALIZATION SYSTEM

The aim of this study is to develop a visualization system that can support the comparison and analysis of the reputations of licensed nursery schools. As an example, in this paper, we conduct a comparative analysis of licensed nursery schools located in Bunkyo Ward, Tokyo, using data provided by the Tokyo Metropolitan Government Bureau of Welfare that includes physical information such as location and capacity [13], description of the philosophy of the nursery school, and a third-party evaluation by the Tokyo Metropolitan Government [3]. The latter data was collected by web scraping.

The visualization system proposed in this paper includes the following three main components.

- Feature classification mapping component.
- Co-occurrence network visualization component for a single nursery school.
- Comparative visualization component of the co-occurrence network for multiple licensed nursery schools.

A. Feature Classification Mapping Component

This visualization component uses third-party evaluation data (from the Tokyo Metropolitan Government in this paper) to cluster the characteristics of licensed nursery schools and display the results on a map, as shown in Fig. 1. The goal of this component is to provide an overview of the characteristics of licensed nursery schools, enabling educational experts to identify groups of schools with similar profiles at a glance. By classifying and visually mapping these characteristics, the component helps users navigate the complex evaluation data and serves as a starting point for deeper analysis.



Fig. 1: Mapping diagram (UMAP, k-means)

To achieve a meaningful classification and visualization of licensed nursery school characteristics, our approach consists of the following three modular steps: (1) feature matrix extraction, (2) dimensionality reduction, and (3) clustering. This modular design follows the common document clustering pipeline structure, as analyzed by Eklund et al. [?], where each step can be independently configured and optimized.

In our current implementation, we employ Sentence-BERT for feature matrix extraction, as our goal is to compare similarities and differences among nursery schools and to identify the underlying factors at the word or phrase level. Sentence-BERT fulfills these requirements by capturing rich semantic representations that allow for fine-grained analysis across textual evaluation data. For dimensionality reduction, we apply PCA and UMAP to compress the feature space to 20 dimensions. For clustering, we use k-means and HDBSCAN to group licensed nursery schools with similar characteristics. Each step is replaceable, allowing flexibility for future extensions. The clustering results are color-coded and visualized on a geographic map, thereby supporting an intuitive understanding of the distribution and characteristics of nursery schools.

We used a Sentence-BERT model [14], an extension of the Japanese pre-trained BERT model developed by Tohoku University [15], for sentence-level feature extraction.

To determine the appropriate dimensionality reduction and clustering methods, we evaluated the consistency between the clustering results visualized in two dimensions using UMAP, as shown in Fig. 2, and the summaries of third-party evaluation data for each nursery school generated by a large language model (LLM). The summarization process involved prompting ChatGPT to produce approximately 200-character summaries that included both “highly evaluated points” and “issues to be addressed.” This paper examines the validity of the clustering results using three nursery schools operated by the same organization: Sakura Mirai Kohinata (hereinafter referred to as “Kohinata”), Sakura Mirai Suido (“Suido”), and Sakura Mirai Gokokuji (“Gokokuji”). As shown in Fig. 2a, Fig. 2c, and Fig. 2d, Kohinata and Suido are grouped into the same cluster, whereas Gokokuji forms a separate cluster. Here, Gokokuji received high evaluations, according to Table I, for its use and operation of manuals, while this aspect was identified as a challenge for both Kohinata and Suido. This discrepancy likely contributed to the observed clustering pattern. Based on these findings, the following combinations of dimensionality reduction and clustering methods were deemed effective for the dataset used in this study: (PCA, k-means), (UMAP, k-means), and (UMAP, HDBSCAN).

However, at this stage, the validation of the consistency between the classification results and the summaries of the third-party evaluations remains subjective. As future work, we plan to incorporate better evaluation methods such as user studies to systematically examine the appropriateness of dimensionality reduction and clustering techniques used in the feature classification process.

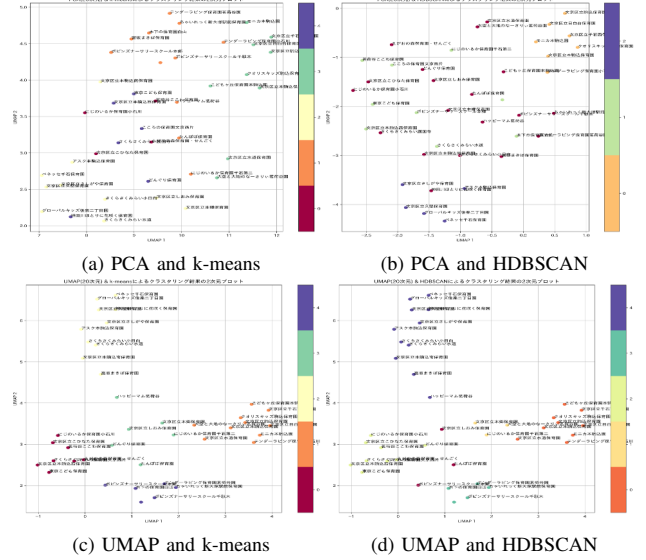


Fig. 2: Comparison of Feature Classification Mapping Methods

B. Co-occurrence Network Visualization Component based on the Data from a Single Nursery School

The purpose of this component is to support a detailed understanding of the characteristics of an individual licensed nursery school based on third-party evaluation data. By constructing a co-occurrence network of frequently mentioned terms, users can intuitively grasp the central topics and their relationships. This visualization also helps identify which topics are positively or negatively evaluated, enabling a more nuanced interpretation of the school’s strengths and challenges. Using the third-party evaluation data of each nursery school, this component generates a co-occurrence network through the following procedure, as shown in Fig. 3.

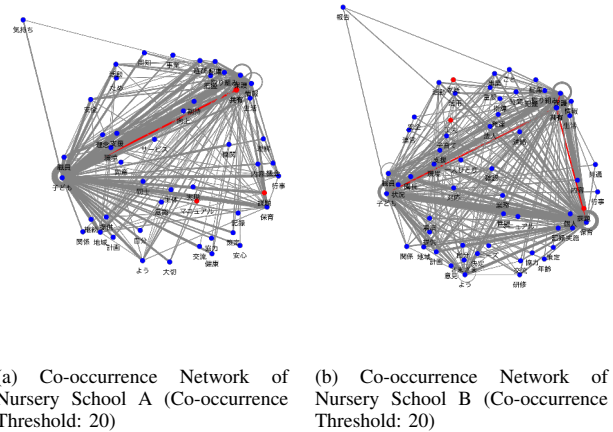


Fig. 3: Co-occurrence Network

TABLE I: Summary of Third-Party Evaluations for Each Nursery School. Texts were partially generated using ChatGPT.

Name of Nursery School	Summary of Evaluations
Sakura Saku Mirai Gokokuji	The institution has established a well-structured system by utilizing guidelines and manuals developed by the corporation, promoting staff training and the standardization of operations. High evaluations were given for practices such as collaboration with guardians and the local community, as well as child-centered childcare. On the other hand, due to the wide scope of activities and parent-related tasks, reducing staff workload and ensuring operational efficiency remain future challenges.
Sakura Saku Mirai Suido	The center received high evaluations for effectively communicating its philosophy to both staff and guardians, implementing planned operations, fostering family cooperation, practicing child-centered care, promoting community interaction, and ensuring safety. On the other hand, strengthening the organizational framework remains an issue, particularly in enhancing knowledge-sharing systems, utilizing near-miss reports, and regularly reviewing manual operations.
Sakura Saku Mirai Kohinata	The center was highly evaluated for its strong implementation in various areas, including the development of a staff system, ongoing internal and external collaboration, thorough infection control measures, respectful relationships with guardians, child-centered care, and attention to safety. Nevertheless, future challenges include revising the BCP (Business Continuity Plan) and operational manuals, ensuring the successful implementation of a new personnel system, and enhancing opportunities for staff reflection to support continuous improvement.

[Step 1:] Extract words from the evaluation data of all licensed nursery schools.

[Step 2:] Select the top k most frequent words from the extracted word set. We set $k = 200$, considering the readability of the resulting co-occurrence network. This number was empirically found to strike a good balance between capturing important topics and maintaining visual clarity.

[Step 3:] Create a co-occurrence matrix (200×200) using the selected words by computing as follows. For each sentence $s_k \in S = \{s_1, s_2, \dots, s_m\}$, we extract a list of nouns and generate all unique unordered word pairs that co-occur within the sentence. The set of co-occurring word pairs for sentence s_k is defined as follows:

$$P_k = \{(w_i, w_j) \mid w_i, w_j \in s_k, i < j\},$$

where w_i and w_j are nouns that appear in the same sentence s_k , and $i < j$ ensures that each pair is considered only once in an unordered fashion. The co-occurrence matrix $\mathbf{C}$ is then constructed by counting the number of sentences in which each word pair appears. Each element C_{ij} of the matrix is defined by

$$C_{ij} = \sum_{k=1}^m \mathbf{1}_{\{(w_i, w_j) \in P_k\}},$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function that returns 1 if the pair (w_i, w_j) is present in P_k , and 0 otherwise.

[Step 4:] Apply Multi-dimensional Scaling (MDS) to reduce the dimensions to two, and determine and save the node positions.

[Step 5:] Extract the words from the evaluation data that correspond to the saved node positions for each nursery school.

[Step 6:] Apply TF-IDF to select the top 100 most important words for each nursery school.

[Step 7:] Create a co-occurrence matrix (100×100) for each nursery school using the selected words.

[Step 8:] Generate the co-occurrence network based on the co-occurrence matrix and the saved node positions.

By fixing the positions of nodes as computed by Multi-Dimensional Scaling (MDS), this system enables comparative analysis of co-occurrence networks constructed from evaluation data of different nursery schools. In addition, the

component attempts to enhance the readability of the network by allowing users to interactively adjust the threshold of edge visibility based on co-occurrence frequency between nodes through a slider. Furthermore, the component applies sentiment analysis to the words represented by each node, with negative nodes colored red and positive nodes colored blue. This color-coding helps education experts, including nursery school operators and local government officials, to support an intuitive and efficient understanding of the evaluation content of each nursery school through the visualization.

C. Comparative Visualization Component of Co-occurrence Network using Data from Multiple Daycare Centers

This component is designed to facilitate the direct comparison of evaluation data between two licensed nursery schools. By visualizing shared and unique co-occurrence relationships within a single network space, users can intuitively identify both similarities and differences in how the schools are evaluated. The aim is to support more informed decision-making by highlighting common strengths, school-specific challenges, and divergent areas of focus. Using the third-party evaluation data from two specific nursery schools, we constructed a co-occurrence network as shown in Fig. 4 by the following steps.

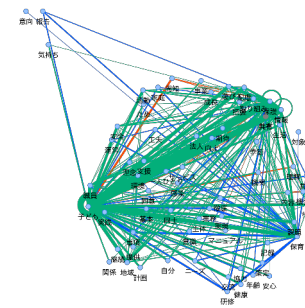


Fig. 4: Co-occurrence Network Constructed from the Evaluation Data of Nursery Schools A and B (Co-occurrence Threshold: 20)

[Step 1:] Extract words from the evaluation data of all licensed nursery schools.

[Step 2:] Select the top 200 most frequent words from the extracted word set.

[Step 3:] Create a co-occurrence matrix (200×200) using the selected words.

[Step 4:] Apply Multi-dimensional Scaling (MDS) to reduce the dimensions to two and treat as the node positions.

[Step 5:] Extract words corresponding to the node positions from the evaluation data of two selected nursery schools.

[Step 6:] Calculate the common and differing words between the two schools.

[Step 7:] Apply TF-IDF to select the top 100 most important words for each school.

[Step 8:] Create a co-occurrence matrix (100×100) for each school using the selected words.

[Step 9:] Generate the co-occurrence network based on the co-occurrence matrix and the saved node positions.

As in the previous section, we fix the positions of the nodes as computed by Multi-Dimensional Scaling (MDS), which enables direct comparison with the co-occurrence network of a single nursery school and ensures a consistent user interface across all network visualizations. Furthermore, we implemented a feature that allows users to switch edge colors based on predefined criteria. In this study, the co-occurrence network shown in Fig. 4 adopts the following color scheme: red edges represent co-occurrence relationships that appear only in “Nursery School A,” blue edges represent those unique to “Nursery School B,” and green edges indicate relationships shared by both schools. By focusing on the green edges, users can visually identify the common features between the two nursery schools within a single, unified network.

IV. EXAMPLE OF THE ANALYSIS RESULTS

A. Feature Classification and Mapping

Regarding the feature classification mapping, Fig. 1 clearly illustrates clusters of nursery schools with similar characteristics. This visualization serves as a valuable starting point for evaluation analysis, as it provides meta-information on which nursery schools share comparable features.

B. Co-occurrence Network Analysis based on Data from a Single Nursery School

We described an overview of nursery schools with similar characteristics in Section IV-A. Then, we focus on detailed analysis and comparison of individual nursery schools using co-occurrence networks constructed from their respective third-party evaluation data.

By examining the number of edges connected to specific nodes, we can visualize and compare the characteristics across multiple nursery schools and identify distinguishing features. Fig. 3a and 3b illustrate co-occurrence networks generated from the third-party evaluation data of “Nursery School A” and “Nursery School B,” which belong to the same cluster. Here, the node labeled “Sharing” is selected in both figures. A comparison reveals that the “Sharing” node in Nursery School B (Fig. 3b) has a greater number of connected edges than that of Nursery School A (Fig. 3a). This suggests that

the evaluations for Nursery School B contain more references related to “Sharing” than those for Nursery School A, thereby highlighting a key difference between the two institutions.

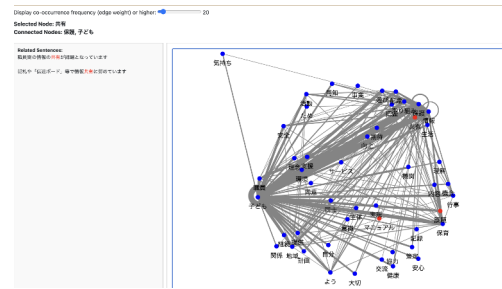


Fig. 5: User Interface for Co-occurrence Network Based on a Single Nursery School’s Data

We also implemented the co-occurrence network component for an individual nursery school as an interactive system, as shown in Fig. 5. This system features a slider for adjusting the threshold of displayed edges and a left-side panel that displays relevant evaluation texts upon node selection. Additionally, the color of each node intuitively indicates the sentiment of the evaluation (positive or negative). For example, in Fig. 6, nodes such as “Sharing,” “Challenges,” and “Manuals” are displayed in red, signifying that these items were subject to negative evaluations. This visual representation enables users to easily identify the concern of the nursery school.

As future work, we plan to conduct user experiments to evaluate the usability and effectiveness of the interface. Additionally, we will consult with educational experts to assess whether the displayed nodes are truly essential for analyzing nursery school evaluations.

C. Comparative Co-occurrence Network Analysis using Data from Two Nursery Schools

As described in Section IV-B, we conducted a comparative analysis of two nursery schools classified into the same cluster to identify and understand the differences between them. In this section, we focus on the similarities between these two nursery schools by analyzing a co-occurrence network shown in Fig. 6 constructed from their third-party evaluation data.

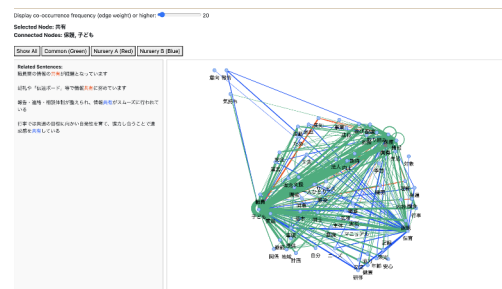


Fig. 6: User Interface for the Co-occurrence Network Based on Data from Two Nursery Schools

We implemented this visualization component for the co-occurrence network with features similar to those described in Section IV-B. It features a slider to adjust the number of displayed edges and a function to display relevant evaluation sentences in the left panel upon node selection. Additionally, users can toggle the edge colors using buttons, allowing them to visualize only the green edges that represent shared characteristics between the two schools.

Figure 6 illustrates the state in which the node “Sharing,” commonly present in both nursery schools, is selected. The system displays the corresponding evaluation sentences with color coding in the left panel, allowing identification of whether each sentence pertains to Nursery School A or B. From this visualization, we can observe that both Nursery School A and B are evaluated from the perspective of “Sharing.” Furthermore, by comparing the contents in the left panel, it is evident that in Nursery School A (indicated in red), the sharing of information among staff members is emphasized, whereas in Nursery School B (indicated in blue), the sharing of information between guardians and staff is highlighted. This comparison clearly reveals differences within the shared perspective. Thus, this visualization enables an intuitive understanding of the aspects and contents that are commonly evaluated between nursery schools, facilitating the identification of similarities and differences.

V. CONCLUSION AND FUTURE WORK

In this study, we developed a visualization system with the ultimate goal of supporting the task of comparative analysis of nursery school evaluations conducted by educational experts, using third-party evaluation data.

For mapping the characteristics of authorized nursery schools, we proposed a method that applies clustering after reducing the dimensionality of context-aware feature matrices. To validate the proposed approach, we utilized ChatGPT to summarize each nursery school’s evaluation data and examined the consistency between these summaries and the clustering results. Based on this comparison, we identified three effective combinations for our dataset: (1) PCA and k-means, (2) UMAP and k-means, and (3) UMAP and HDBSCAN.

We also incorporated an analysis using co-occurrence network visualizations. This feature enabled not only detailed interpretation of the clustering results but also more in-depth analysis of individual nursery schools and comparisons highlighting similarities and differences with other nurseries.

Future work includes determining the optimal method for feature classification. At present, the evaluation of clustering consistency with the summarized data relies on subjective determination. Therefore, we plan to conduct user studies to assess the validity of the clustering results. In addition, with regard to the co-occurrence network-based analysis, we aim to evaluate whether the current system meets the analytical needs of educational experts examining regional early childhood education. To achieve this goal, we plan to conduct usability testing targeting users with expertise in education.

We showcased the use of our system with data from licensed nursery schools in Bunkyo City; however, its design allows it to be applied to similar data from other municipalities.

ACKNOWLEDGMENT

Prof. Akemi Miyazato (Institute for Human Life Innovation, Ochanomizu University) and Prof. Machiko Tsujitani (Faculty of Letters and Education, Ochanomizu University) are gratefully acknowledged for their invaluable advice and support.

REFERENCES

- [1] Ministry of Education, Culture, Sports, Science and Technology (MEXT), “Future directions of elementary school education (report),” https://www.mext.go.jp/b_menu/shingi/chousa/shotou/049/shiryo/icsFiles/afieldfile/2016/02/08/1363649_001.pdf, Feb. 2016, document distributed at the 49th Meeting of the Subcommittee on Elementary and Secondary Education.
- [2] Children and Families Agency (CFA), “Childcare policy guidelines (separated edition: Section 1–24),” https://www.cfa.go.jp/assets/contents/node/basic_page/field_ref_resources/eb316dce-fa78-48b4-90cc-da85228387c2/f4758db1/20231013-policies-hoiku-shishin-h30-bunkatsu-1_24.pdf, Oct. 2023, (in Japanese).
- [3] Tokyo Metropolitan Government, “Tokyo social welfare navi: Third-party evaluation system,” <https://www.fukunavi.or.jp/fukunavi/hyoka/hyokatop.htm>, (in Japanese).
- [4] L. Hirsch and S. Andrews, “Visualising text co-occurrence networks,” in *International Conference on Conceptual Structures*, Jul. 2016.
- [5] M. Grootendorst, “Bertopic: Neural topic modeling with a class-based tf-idf procedure,” *arXiv preprint*, Mar. 2022. [Online]. Available: <https://arxiv.org/abs/2203.05794>
- [6] A. Eklund, M. Forsman, and F. Drewes, “An empirical configuration study of a common document clustering pipeline,” *Northern European Journal of Language Technology*, vol. 9, no. 1, 2023.
- [7] L. Zhang, Q. Jiang, W. Xiong, and W. Zhao, “Evaluating the quality of digital education resources based on learners’ online reviews through topic modeling and opinion mining,” *Education and Information Technologies*, vol. 30, 2025.
- [8] A. K. Jayaraman and A. S. Murugappan, “Opinion-based co-occurrence network for identifying the most influential product features,” *Journal of Engineering Research*, vol. 8, no. 4, Nov. 2020.
- [9] M. Kim, E. Kim, and K. Kinjo, “Comparison of japan and south korea regarding guidance course and contents for students with severe and multiple disabilities in special needs school,” *Journal of Inclusive Education*, vol. 13, pp. 83–104, Aug. 2024, (in Japanese).
- [10] D. Ip, K. K. K. Lau, W. Cui, H. Qu, and H. Shen, “A visual approach to text corpora comparison,” in *International Workshop on Visual Analytics in conjunction with IUI (IVITA)*, Hong Kong, Feb. 2010.
- [11] R. Sevastjanova, S. Vogelbacher, A. Spitz, D. Keim, and M. El-Assady, “Visual comparison of text sequences generated by large language models,” in *IEEE VIS Workshop on Visualization for Data Science (VDS)*, Oct. 2023.
- [12] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts, “Visual comparison for information visualization,” <https://graphics.cs.wisc.edu/Papers/2011/Gle11>, Aug. 2011, preprint.
- [13] Tokyo Metropolitan Government Bureau of Social Welfare and Public Health, “List of certified nursery schools in tokyo (type a and b),” <https://www.fukushi.metro.tokyo.lg.jp/kodomo/hoiku/ninsyo/ichiran.html>, (in Japanese).
- [14] I. Sonobe, “Japanese sentence-bert model v1,” <https://huggingface.co/sonoisa/sentence-bert-base-ja-mean-tokens>, 2020.
- [15] Tohoku University NLP Group, “Pretrained japanese bert models,” <https://github.com/cl-tohoku/bert-japanese>, 2019.