2O1-ES-5-02

Grasping Research Trends Based on Similar Cited Papers

Shun Gendo*1 Koichi Hodooka*1 Mitsunori Matsushita*1

*1 Kansai University

The goal of this study is to develop a system that supports grasping trends of the research field by visualizing among research papers.Previous studies have visualized the relations between sources and citations to understand trends in research fields.However, these methods only focus on a single paper, and it is difficult to find relations when there is no explicit reference, even though the papers are close in content.In order to solve these problems, this paper developed a prototype interface that helps to grasp relations between similar sources and citations over papers.

1. Introduction

In recent years, many papers have been written in various fields[1]. Researchers read dissertations in their own field of research to determine their research position. When a researcher reads a paper, they understand the relationship between a dissertation that is compared with another dissertation and a dissertation that forms a technical basis. Through this process, researchers can observe research relationships and understand research trends.

Relationships between papers can include, for example, similar backgrounds or similar technologies being used in the proposed method. Such relationships help researchers decide to collect the dissertation, and they are not explicitly stated. Instead, researchers must enter keywords and author names in search engines(e.g., GoogleScholar, CiNii) to find papers.

However, searching using keywords will not yield all papers in the same research field. Moreover, searches using author names will not provide only papers in the research area for which they are intended. Researchers will not accept satisfactory results for keyword or author name searches. However, researchers can obtain papers by searching the cited papers listed at the end of a paper.

Thus, this study focused on cited papers. By visualizing the citation relationships of papers, this paper aims to support the understanding of research trends. As a starting point, the network will consider an interface that supports the comparison of cited papers.

2. Related works

To efficient the process of surveying research papers, a method of visualizing citation relations and cited relationships in papers has been proposed.

Inoue et al. proposed a system for visualizing citation relationships and cited relationships between papers[2]. The visualization system uses each paper as a node and expresses the citation relations between papers as edges. However, displaying all the citation relationships and cited relationships of papers would increase the number of papers. As the number of papers increases, their appearance becomes disordered. To solve this problem, Inoue er al. has built a function that filters the display of papers for the reasons cited. By using the filtering function, the efficiency with which users could search for papers was improved. However, in this system, because the visualization screen and the reason for citation are displayed on separate screens, the user will struggle to grasp the relationship between papers when only using the visualization screen.

Maki et al. mentioned that natural language has spelling variations, which reduce the accuracy of searches and analysis[3]. To solve this problem, they proposed a dictionary called Broadic, which compensates for variability in writing. Broadic was built using I-Scover (a paper search system of the Institute of Electronics, Information and Communication Engineers including technical terms) and Wikipedia. I-Scover is a paper retrieval system made by the Institute of Electronics, Information and Communication Engineers (IEICE). A visualization system that analyzes the technical factors arranged in chronological order using the data of papers in I-Scover was proposed. When using this system, they evaluated the effect of visualization with and without Broadic. Consequently, they could visualize more accurate research by trends using Broadic. However, to successfully grasp the research trends of a user, many papers must be collected.

Research has been conducted to predict the central researcher in a sprouting field by using a coauthor network. Mori et al. considered authors with a high number of citations as the main investigators in their research area[4]. Then, to identify the research area, a co-author network was obtained for the original researcher. A learning model to predict the number of citations of researchers in a certain research area was constructed based on the features of the network structure. This learning model was used to make predictions for central researchers. However, it is difficult to grasp whether the paper of the central researcher belongs to research field intended by the user. Furthermore, it is difficult to grasp the differences between papers.

3. Position of this study

Research that visualizes citation relationships and cited relationships among papers will aid understanding of the

Contact: Shun Gendo, Faculty of Informatics, Kansai University, Ryozenji, Takatsuki, Ohsaka 569-1095, Japan, k911867@kansai-u.ac.jp

relationships between citation-related papers. However, it is difficult for users to understand the relationships between uncited papers.

Research that promotes understanding of research trends will enable the analysis of trends in each research field. However, it is difficult to understand the differences in the content of the research. Thus, it is difficult to understand what research activities are occurring within the same research field.

In studies that analyze author information, the use of co-authorship can lead to a more centralized author. However, it is difficult to grasp research trends without understanding the research conducted by highly central authors. In addition, authors with high centrality may not necessarily be important authors for the user's research area.

This study focuses on cited papers. Cited papers are often used to analyze the relationships between papers. By using cited papers, we can visualize the research trends of the author of the paper selected by the user. This will help in grasping papers that are unrelated to the citations. Other papers by the authors of the papers selected by the user could be outside the research field intended by the user. We want to allow users to find papers with high similarity to the paper selected by the user while understanding the differences between papers with high similarity. This study aims to realize an interface that supports the identification of research trends.

4. Design guidelines

When comparing multiple papers, cited papers can help identify other papers in the same field, while also helping promote understanding of the relationships between papers. This paper focuses on the paper to be compared (hereafter, the comparative paper)with the paper that the researcher wants to read (hereafter, the starting paper). Cited papers that are common to the originating paper and the comparative paper can indicate the identity and relevance of the research field. Further, the differences between the viewpoints and emphasis between the papers can be grasped from the differences between the cited papers of the original paper and the comparative paper. Understanding the differences between the focus and emphasis requires an understanding of the position of the research.

Papers by the same author are often papers in the same field. Viewing the same author's papers over time helps to keep track of research trends. Papers often involve more than one author. Authors must be able to view their papers. Considering the above-mentioned points, the three design guidelines are as follows: 1. To be able to grasp the degree of coincidence of cited papers 2. To facilitate identification of the difference between cited papers 3. View papers by each author

5. Proposed method

We constructed a database of 350 papers in the field of comic engineering. This database comprised paper collections, author names, publication years, and cited papers. Using this database, a research trend grasp support interface was implemented (see Figure 1) using HTML, CSS, JavaScript, and jQuery. The visible part used Cytoscape.js, which is an open-source JavaScript-based graph library provided by Franz et al. Cytoscape.js is a visualization software component, which can render interactive graphs in a web browser[5]. Using these, we implemented an interface to facilitate the comparison of cited papers.

Graph drawing is useful for intuitively expressing human relationships and citation relationships in papers[6]. To achieve this, the proposed interface is drawn using nodes and edges. The nodes of the proposed interface represent papers, whereas the edges of the proposed interface represent citation relationships and cited relationships. Nodes that have citation relationships and cited relationships with the author's paper are connected by edges. The screenshot in Figure 1 shows the research trend of the author (originating author) who wrote the paper (originating paper) selected by the user. The red nodes represent the originating paper. The blue nodes represent other papers written by the original author (see 1 in Figure 1). The orange and gray nodes below these nodes (see 2 in Figure 1) are lists of cited papers. In the papers by the original author, the citation-related papers are indicated by edges (see 3 in Figure 1). The triangular arrow denotes the quotation destination, and the circle mark denotes the quotation source. Among the edges, the black edge indicates a citation related to the originating paper, whereas gray edges indicate citation relationships and cited relationships other than the originating paper. Each node is displayed in chronological order. Consequently, it is possible to intuitively understand which papers are cited sources and which papers are cited by the coordinates at which nodes are displayed.

6. Something similar

To satisfy Design Guidelines 1, the nodes of the cited papers common to the origin paper and the comparative paper are displayed in orange. In addition, different nodes are individually displayed in gray (see 2 in Figure 1). By using colors for the nodes, it is possible to intuitively understand the degree of coincidence between the cited paper of the starting paper and that of the comparative paper. For example, suppose there is a comparative paper with many orange nodes in the list of cited papers. It can be inferred that this comparative paper has high similarity with the starting paper and the purpose and method. Conversely, for a comparative paper with many gray nodes in the list of cited papers, it can be inferred that the paper has low similarity with the starting paper and the purpose and method. Furthermore, if there are no comparative papers that match the list of cited papers in the originating paper, it can be inferred that the research direction of the originating paper has changed. In addition, if there are multiple comparative papers that match the list of cited papers of the original paper, it can be understood that there are multiple studies with high similarity in terms of the purpose and method of the original paper When the user clicks the cited paper

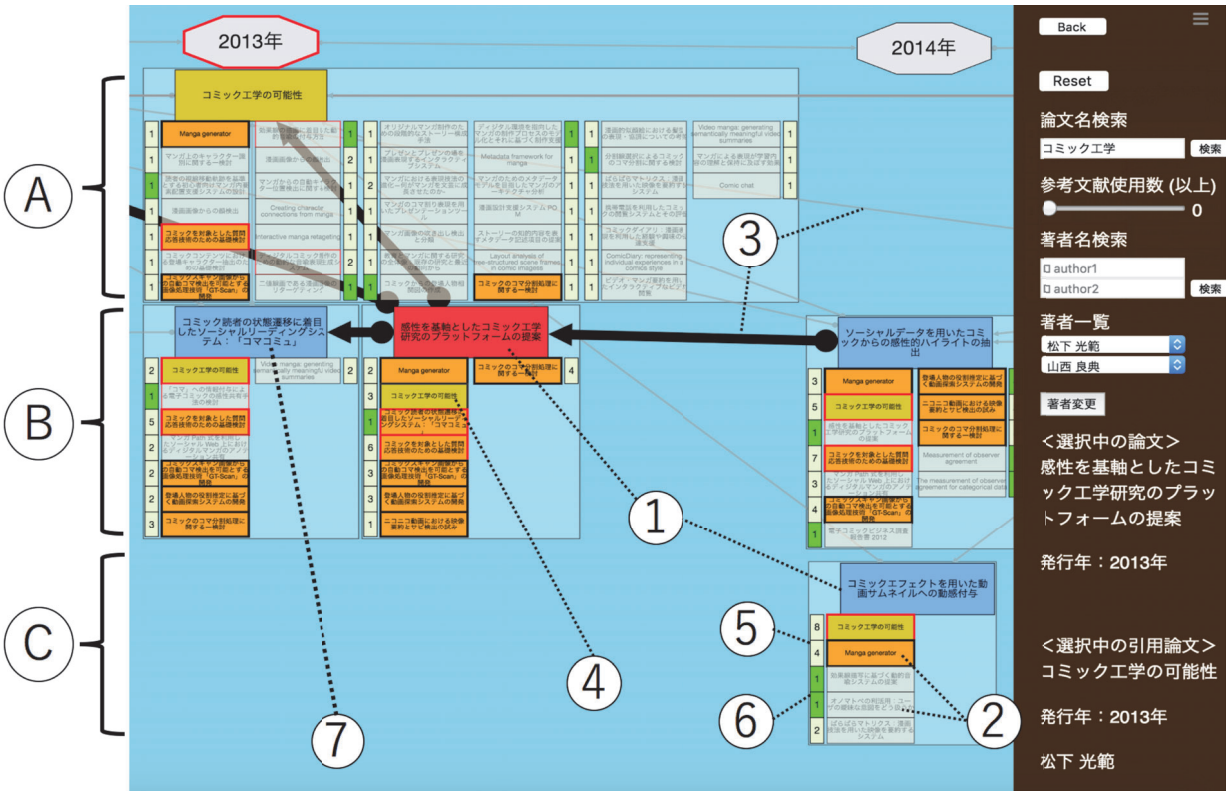

Figure 1: Proposal interface

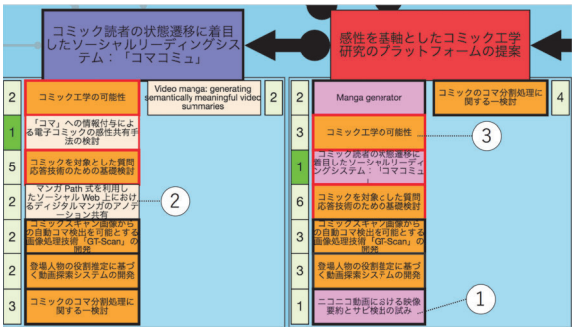

Figure 2: Changes in cited papers

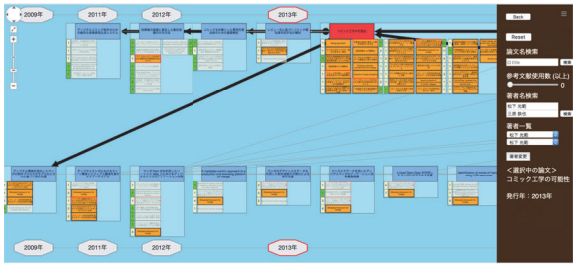

Figure 3: Research trends among authors who are not co-authored

node, the node of the cited paper that is the same title as the clicked cited paper turns yellow (see 4 in Figure 1). By observing the nodes that have changed to yellow, it is possible to confirm to what extent the cited paper was cited and in what papers. The numbers displayed next to the nodes of cited papers (see 5 and 6 in Figure 1) indicate the number of times the origin author cited the cited paper. If the number is 1, the author has cited the cited paper only once. The cited paper was not cited before the year the cited paper was cited. Therefore, the user does not need to check whether the cited paper is cited in a paper that is a year older than the original paper. If the background of the number is dark green (see 6 in Figure 1), it indicates that the cited paper has not been cited since the original paper. Therefore, when searching for a paper citing the paper, the user does not need to refer to a newer year than the paper citing the paper. If the node is enumerated, the number is 1, and the node is dark green, it can be inferred that the cited paper is cited only by the citing paper. Thus, the burden of the user moving the screen has been reduced.

Next, to satisfy Design Guideline 2, node colors are distinguished between cited papers that are common between the two papers and papers that are cited only in one paper. Right-clicking on a paper other than the original paper (see 7 in Figure 1) changes the color of that paper and the papers cited in the original paper (see Figure 2) The pink nodes (see 1 in Figure 2) indicate papers cited only in the source paper. The yellow-white nodes (see 2 in Figure 2) denote papers cited only in newly selected papers. The cited paper that is common to the two papers is indicated by the orange node.) The differences between the papers can be intuitively understood by these color changes. A cited paper that appears in only one paper is crucial to understanding

the characteristics of the paper that cited it.

To satisfy Design Guideline 3, the author's name is displayed at the y-coordinate. A paper written by the first author is shown in the upper row (see A in Figure 1). A paper written by the second author is shown in the bottom row (see C in Figure 1). A paper co-authored by both parties is shown in the middle (see Figure 1B). In this way, the user can easily understand which paper is written by each author. Further, the user can understand the relationship between the authors. Figure 3 shows the research trends of two authors arbitrarily selected by the user. This display allows a comparison of research trends among authors who are not co-authored. Consequently, the user can understand the difference between the research fields and the cited papers.

7. Conclusion

This paper proposed an interface to understand the relationships between papers in order to grasp research trends. The purpose of the interface implementation is to visualize the differences between cited papers, the corresponding cited papers and paper's authors. In future, the usefulness of the interface for the lead users should be verified.

References

[1] Isaka, N. and Nakayama, T: Iask: A Survey System for Searching Important Articles, *IPSJ SIG Technical Reports*, Vol. 2011- CE-109 (2011).

[2] Han, D. and Inoue, A.: Efficiency improvement of literature survey based on citation-reason visualization, *Institute of Electrical and Electronics Engineers*, pp. 189-193 (2017).

[3] Maki, T. and Wakahara, T.: Time Series and Technology Factor Analysis Using I-Scover Literature Metadata, *The journal of the Institute of Electronics, Information and Communication Engineers*, Vol. 99, No. 10, pp. 1002-1012 (2016).

[4] Mori, J., Hara, T., Sakaki, T., Kajiyama, Y. and Sakata, I.: Predicting Citations using Citation Network and Text Analysis from the Academic Paper Database, *The 29th Annual Conference of Japanese Society for Artificial Intelligence*, Vol. JSAI2015, pp. 1B21-1B21 (2015).

[5] Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O. and Bader, G. D.: Cytoscape. js: a graph theory library for visualisation and analysis, *Bioinformatics*, Vol. 32, No. 2, pp. 309-311 (2015).

[6] Toeda, N., Nakazawa, R., Itoh, T., Saito, T. and Archambault, D. W.: On Edge Bundling and Node Layout for Mutually Connected Directed Graphs, *Information Visualisation (IV), 2016 20th International Conference*, IEEE, pp. 94-99 (2016).