# A Paper-Search System to Identify Citation Networks and Motivations

Taerin Yoon, Hyunwoo Han, Hyoji Ha*
Lifemedia Interdisciplinary
Program, Ajou University

Juwon Hong, Kyungwon Lee†
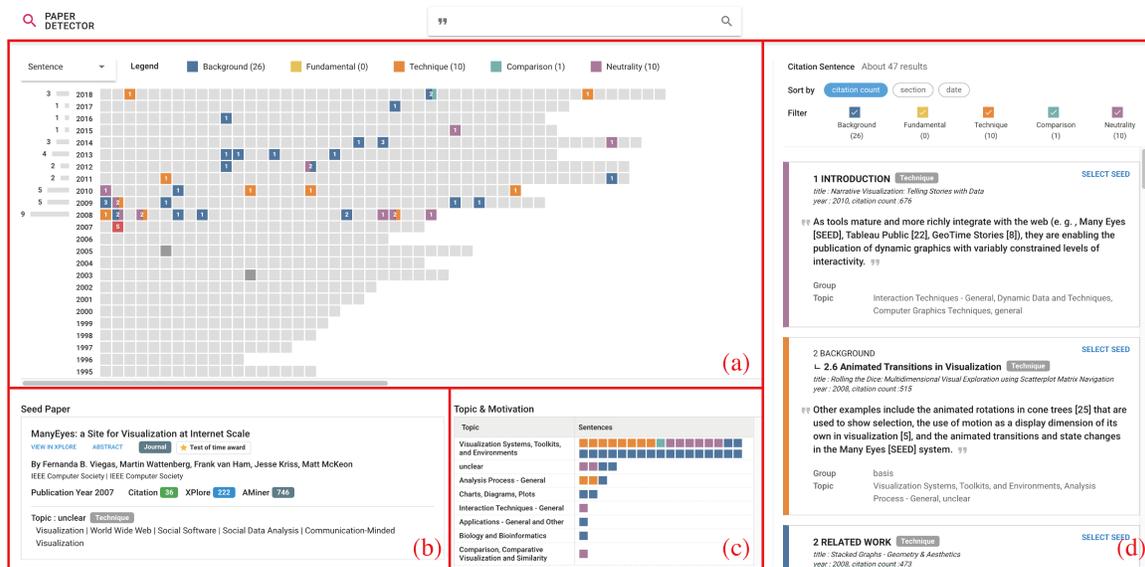Department of Digital Media,
Ajou University

Figure 1: Overview of the system to identify motivations for paper citations. (a) Citation network component view: every node represents a paper and the number within the node indicates the number of times the target paper is mentioned. The color of the node stands for different motivations for citation. Also, papers cited by the target paper are shown in dark gray. (b) Card view providing detailed information on the paper, (c) topic and motive status of a cited sentence and (d) detailed information list on the cited sentence. Interface when *Many Eyes* is selected as the target paper (red nodes "S").

## ABSTRACT

We present a system that can determine the motivation and topic of citations when a user is searching for appropriate reference literature. To build the system, we extracted citation sentences from papers from the IEEE Information Visualization Conference and analyzed the motives for citing specific papers. We also generalized the topics by extracting keywords from the sentences. With this motive and topic data, a paper-search interface service was designed. Finally, we checked the motive and topic information for a cited sentence by using our system on a user scenario.

**Keywords:** Scholarly data visualization, citation motivation, topic analysis, citation network.

**Index Terms:** Information interfaces and presentation—User Interfaces—Graphical user interfaces

## 1 INTRODUCTION

Researchers gather and analyze data that could potentially be cited to make a convincing argument about their research and use various academic motivations to select documents to cite. However, as the already enormous amount of literature continues to increase, the scale of the expanded research publications makes an in-depth literature search difficult.

*e-mail: {yoontr, ainatsumi, hjha0508}@ajou.ac.kr
†e-mail: {penguin373, kwlee}@ajou.ac.kr

If a researcher with a specific motive decides to cite a particular paper and another paper has cited the identical paper with the same motive, the paper can be interpreted as being worth reading. Consequently, this study proposes a system that can determine citation motives and topics from a cited sentence extracted from an IEEE InfoVis paper and visualize the results. When a user searches the literature using our system, our aim is to help them determine how to cite the target paper in their paper and judge its value. To achieve our research goal, the following research process was carried out.

1) We collected previous studies that have suggested motives for citation and based on collected data, citation motive types were defined.

2) We extracted keywords from the cited sentences by implementing syntactic and entity analyses and generalized the study by matching the keywords with visualization-field research topics.

3) Based on the data about citation topics and motives, we designed a paper-search interface system.

4) Using a scenario, we explained how a user can search for citation motives and determine topics.

By examining the research topics and motives of papers citing the target paper, we can observe how the target paper was developed and reproduced by other papers.

## 2 DATA COLLECTION AND REFINEMENT

### 2.1 Citation Keyword and Topic Extraction

In this research, texts are obtained from PDF files from InfoVis conference papers and the cited sentences are extracted by referring to the citation characteristics and reference methods of IEEE InfoVis papers. Next, we used "entity analysis" and "syntax analysis" from

the Google Cloud Natural Language API to extract the keywords and topics [2]. The keywords are clustered into topics using KeyVis data, which proposes concept maps for the diverse terms used in the Visualization research field and the sentences are classified using the clustered topics.

## 2.2 Citation-Motive Classification

In this study, the citation motivations are divided into five classification classes, referring to Dong et al. [1] and Taufel et al. [4]. We used four of the Background, Fundamental, Technique, Comparison proposed in the study of Dong and Neutral as the last one from Taufel's research. Dong defined four types and proposed cue words that can classify them. Therefore, in this study, we gathered words and their synonyms based on Dong's study and constructed a cue word dictionary with cue words found during the study. The five categorized motives classified through the presence or absence of cue words are as follows.

- **Background**: Citation that describes the background of a complete topic or provides a recent study or cutting-edge approach for a general method.

- **Fundamental**: Citations of major previous articles that inspire or give concrete hints for current works.

- **Technique**: Citations of important tools, methods, data and other resources used/recomposed in the current study.

- **Comparison**: Citations comparing methods or results with current research.

- **Neutral**: A paper that exists in the bibliography but does not appear in a substantial sentence. In other words, the related text information cannot be checked in the text.

## 3 SYSTEM INTERFACE AND INTERACTION

The system interface of this study consists of four parts, as shown in Figure 1. Each view is driven by selecting the target paper that the user wants to figure out and the target paper can be selected through a citation-network view or by searching the title of the papers. Figure 1(a) is the citation-network component view, which shows the properties of a paper node. This view offers a mode for showing the sentence-based properties of a paper node. When this mode is selected, every paper node that cites the target paper is activated [3]. The activated nodes are colored according to their citation-motive classification, showing how many times the target paper appears in the cited paper.

Figure 1(b) presents a view that shows detailed information about a specific paper node when it is hovered over. Figure 1(c) is a topic-and-motivation view that shows the status of the motives and the topics of the citations. The provided topic information is placed in the first column of the table; each node that appears in the second column represents a single citation and is colored according to its citation-motive classification.

Figure 1(d) is a view that allows the user to check a detailed list of citations from the paper that is citing the target paper. Each card view in the list represents a citation, with the original sentence text, the title of the paper and the heading of the section where the sentence appeared in the paper. In addition, the user can confirm a topic without reading the sentence firsthand because the color corresponding to the citation motive is provided, along with the extracted study-topic data from the sentence keyword.

## 4 USER SCENARIO

In this section, we present a user scenario to describe how someone would use our system to search citation motives and find a topic. Let us suppose that a particular user was inspired by the *Many Eyes* study [5] announced in IEEE InfoVis 2007 and uses our system to find the study cited with a *'Technique'* motivation. The user sets the *Many Eyes* paper as the target paper by selecting the row for the year the paper was published. Figure 1 shows the interface with *Many*
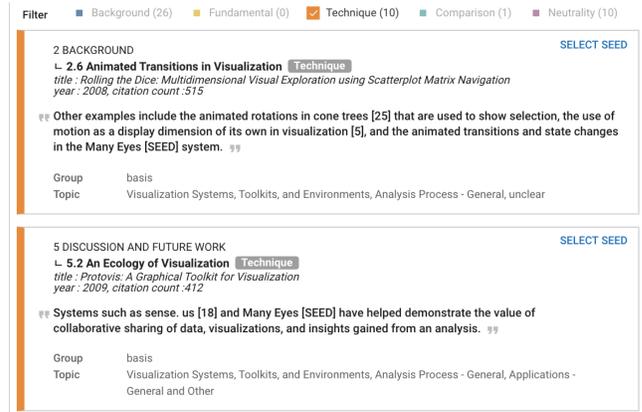


Figure 2: Filtering is applied in the Citation sentence detail view only to check sentences classified as *'Technique'*. The 'SELECT SEED' button can be used to select the paper of the sentence as the target paper.

*Eyes* selected. From the results, we can verify that *Many Eyes* was cited nine times in 2008, the year after it was announced. In fact, 30% of the papers that year cited it. With the network component view, it is possible to confirm that the paper was cited continuously until 2018, about 11 years later. *Many Eyes* is cited in the order of *'Background'* (navy), *'Technique'* (orange) and *'Neutral'* (purple) in sentences where 'Visualization Systems, Toolkits and Environments' is the citation topic.

At this point, the user selects the *'Technique'* motivation through the Filter function in the view, which allows the user to find only research based on technical motivation. (Figure 2) This helps identify sentences in other papers related to technical motivations.

## 5 CONCLUSION

This paper proposed a paper-search system using motives and paper topics. It allows a user to investigate how including a target paper in their own research when seeking for literature; it also helps users judge the value of the target paper. In future work, we plan to improve the algorithms to more precisely extract citations and keywords. In addition, the paper-search system will add a mode that shows dissertation-node properties based on the paper topic. Furthermore, the systematic usability of the system will be verified.

### REFERENCES

[1] C. Dong and U. Schäfer. Ensemble-style self-training on citation classification. In *Proceedings of 5th international joint conference on natural language processing*, pp. 623–631, 2011.

[2] Google Cloud. *Cloud Natural Language — Cloud Natural Language API*, January 2019.

[3] J. Stasko, J. Choo, Y. Han, M. Hu, H. Pileggi, R. Sadana, and C. D. Stolper. Citevis: Exploring conference paper citation data visually. *Posters of IEEE InfoVis*, 2, 2013.

[4] S. Teufel, A. Siddharthan, and D. Tidhar. Automatic classification of citation function. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pp. 103–110. Association for Computational Linguistics, 2006.

[5] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE transactions on visualization and computer graphics*, 13(6):1121–1128, 2007.