

# A Conference Paper Exploring System Based on Citing Motivation and Topic

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

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

## ABSTRACT

Understanding and maintaining the intended meaning of original text used for citations is essential for unbiased and accurate scholarly work. To this end, this study aims to provide a visual system for exploring the citation relationships and motivations for citations within papers. For this purpose, papers from the IEEE Information Visualization Conference that introduce research on data visualization were collected, and based on the internal citation relationships, citation sentences were extracted and the text were analyzed. In addition, a visualization interface was provided to identify the citation relationships, citation pattern information, and citing motivation. Lastly, the pattern analysis of the citation relationships along with the citing motivation and topic was demonstrated through a case study. Our paper exploring system can confirm the purpose of specific papers being cited by other authors. Furthermore, the findings can help identify the characteristics of related studies based on the target papers.

**Keywords:** Citing motivation, topic analysis, citation network, data visualization, conference paper exploring.

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

## 1 INTRODUCTION

In academic papers, “citation” is a form of argument often used to supplement the author’s logic or to justify a claim [11]. Researchers search for previous studies that may serve as a valid source for their research and have a motivation for citing the paper within the citation process of preceding literature. Motivation is a significant element because it is able to understand the citer’s intention. After finding the intention of the citer, we can figure out the inherent meanings, such as whether the current paper is based on previous work, or whether the citer is criticizing particular articles - and indeed, what is the subject of criticism -. However, the volume of literature continuously increases over time, making it difficult for researchers to explore literature in-depth. Accordingly, researchers go through a citation chaining process in which they select several papers and then track and search the references in those papers. They then find the context mentioned in the cited document and directly read it and evaluate it. In addition, the researchers read the context in the cited literature directly and subjectively interpret the author’s motivation for citation contained in the sentence. Here, it is possible to interpret the author’s motivation for citation differently than intended. To address this problem, this study focuses on what additional information, if any, could help researchers search more efficiently when exploring literature. If information on the citer’s motivation were to be provided in the retrieval system, so that the researcher had no need to subjectively interpret the citer’s motivations, they could

\*e-mail: {yoont, ainatsumi, hjha0508}@ajou.ac.kr

†e-mail: {penguin373, kwlee}@ajou.ac.kr

reduce the time needed for the selection process, and avoid any inappropriate alteration of the originally intended meaning. Furthermore, the research topic of the paper citing a specific target paper, and the citing motivation information, can both be used to identify how a specific document is reproduced in the citing paper.

Under this premise, this study aims to find the citing motivations and topics of citations from citations extracted from papers presented at the IEEE Information Visualization Conference (hereafter referred to as “InfoVis”) and apply the results to actual literature exploration. The research process was as follows.

First, we identified previous studies that proposed citing motivations and defined the types of citing motivations appropriate for this study. Second, to analyze the motivations of the papers which were cited in the literature, the lexical and syntactic features were found and defined as the classification criteria, and an analysis was performed accordingly. Third, we extracted the core text from a PDF file and searched for the citations in which the specific paper appeared. Fourth, the syntax and entities of the citations were analyzed to find keywords in the sentences and to generalize them through matching with the research topics used in the visualization field. Fifth, an interface was designed to present the found citing motivation and topic information from the paper search process. Sixth, we introduced how the designed interface interacts and describes the characteristics of the visualization results through a case study. This study was first introduced through a poster session at IEEE VIS2019 [17] and presents a detailed process of how to extract the citing motivations along with a case study to explore the citation relationships of a paper. Specifically, the primary contributions of our work are:

- We presented an overview system based on tile-based visualization to confirm the overall citation relationship and pattern of the conference papers. Our visualization method has the advantage of showing the relationship about the cited or citing papers corresponding to the selected paper. Also, the users can easily confirm the research topic and distribution of citing motivation for specific papers.
- Our system not only provides the title and meta information of paper citing the target paper but also makes user to see the actual sentence of where cited papers are mentioned.

## 2 RELATED WORK

As citations in papers have more complex meanings than simple references which only express one aspect of the original use, researchers have long studied citations across various disciplines. Traditional citation analysis studies have focused on finding influential papers using statistical analysis or revealing the relationships between the literature and authors. There have been recent cases using citation networks to observe which papers are cited the most and which papers have the greatest influence.

Choi et al. [2] conducted a network visualization study to confirm the academic genealogy through an analysis of citation relationships. The advantage of this system is that it can quickly identify academic groups and influential papers based on the citation relationships. However, one limitation is that it does not take into account semantics such as the citer’s motivation.

Next, Stasko et al. [14] presented a CiteVis visualization that shows the relationship between the citation count and citations using papers published in InfoVis. In this study, the papers are indicated by circular nodes, and the citations counts are indicated by brightness. Additionally, each paper was arranged according to the row representing the year of publication, and the nodes were sorted in descending order of citation count. However, this study cannot confirm how the actual cited sentence was made, and does not provide information on the motivation for the citation of the sentence.

Researchers have long investigated citing motivations to determine the fundamental reason for a researcher to cite a particular paper [4, 13]. There has been recent research on automated citation analysis, which automatically finds the meaning and motivation inherent in citation by combining text analysis and machine learning using natural language processing (NLP). Dong et al. [5] conducted a study to analyze the citing motivations in citation sentences using machine learning. The researchers created an ensemble learning model and performed classification considering not only the lexical features but also the physical location of the citations and the syntax structure of the sentences. Because researchers' motivations for citing other papers can be interpreted in various ways, this study also required classification criteria including motivations for general citations. Accordingly, this study used a total of five classification criteria, adopting the four types of citation purposes proposed in the study of Dong and a class that judges the paper as 'Neutral'. Section 3.3 covers this in more detail. In addition to check the citation relationship and the number of citations, the original text of the citation was also provided, along with information on the motivation for citation.

### 3 DATA COLLECTION AND REFINEMENT

#### 3.1 Metadata and KeyVis data collection

First, we collected 718 Papers published in InfoVis from 1995 (in which the InfoVis conference began) to 2018 in PDF files. Next, this study also utilized information from vispubdata.org, which was devised in the study of Isenberg et al. in order to collect the citation literature metadata. [7, 8]. Papers published in InfoVis must include 'Index Terms', the keywords that authors assign to their papers, which are meaningful because these reflect what the authors consider important. Therefore, the first keyword appearing in the 'Index Terms' of the paper was set as the research topic that best describes the paper. Research by Isenberg et al. [9, 10] enabled the KeyVis data of keyword from the 'Index Terms' of 2,433 papers published in the VIS Conference (InfoVis, SciVis, VAST) from 1990 to 2015 to be found on the website, which was used in this study.

#### 3.2 Citation keyword extraction

In order to extract information about cited paper from the citing paper, we used a text similarity measure based on the Dice's coefficient method [3]. Text similarity based on Dice's coefficient is an algorithm that measures how common two strings compared with the number of bigrams, where bigram means adjacent character pairs in the string. For applying the algorithm, we converted the title string ( $x$ ) of the target article and the text ( $y$ ) of the bibliography to lowercase and checked the text similarity. The formula is as follows:

$$d(\text{Dice's coefficient}) = \frac{2n_t}{n_x + n_y}$$

$n_t$ : the number of character bigrams found in both strings

$n_x$ : the number of bigrams in string  $x$

$n_y$ : the number of bigrams in string  $y$

Through the formula, we found the reference number of the target paper with the highest similarity by traversing the list of all references, and then extracted the citation sentences in which the reference number appears. In addition, we got the subtitle of the section and higher-level section included the citation sentences. Table 1 is the list of the citation sentences extracted through the above process.

Table 1: Information about cited paper extracted from the citing paper.

Entity	Description
Sentence	Citation of the target paper from the citing paper
Context	Context including the sentences before and after the sentence citing the target paper in the citing paper
Section	Subtitle of the section to which the citation belongs
Section tree	Higher-level section of the section to which the citation belongs
Reference number	Reference number assigned in the bibliography by the author

Also, we used text analysis to find the keywords and then abstracted these keywords into topics. These topics provided information that could be used to determine what concepts of the visualization research field the citing sentence used to describe the cited paper. Next, to analyze the text, this study used the entity analysis and syntax analysis from the Cloud Natural Language API [6] provided by Google.

Entity analysis in the Google Cloud Natural Language API examines known objects such as celebrities and landmarks in a given text and returns information about the entity. However, when extracting keywords in the form of compound words, due to the limitation of not being able to extract them in the perfect form, a syntax analysis was used to analyze the structure and dependencies of each word in a sentence.

#### 3.3 Citing motivation analysis

As such, they were described as useful for building automatic citation classification models. Therefore, this study adopted the four citation classifications proposed in that study. However, as there were papers in the bibliography that did not appear in the text, there was a need to define additional new classes to classify these cases. With reference to Taufel et al. [15], this study added the 'Neutral' class for cases in which where there is insufficient evidence to classify the motivation type, or for when a paper is referenced but not cited in the main body of text. The descriptions for each motivation 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 implications for current works.
- **Technique:** Citations of important tools, methods, data and other resources used or reapplied 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.

Next, a word dictionary was constructed to detect the citing motivation of the words extracted from the citations of the paper. Cue words were collected for each type of citation purpose defined in



Figure 1: Overview of this study's system for identifying motivations for paper citations. (A) Network overview for viewing citation patterns and motivation information: (a1) Dropdown view for selecting overview options (has sentence base and paper base mode) (a2) Bar graph area showing number of citing papers by year (a3) Citation network view that sets the attributes of the paper node based on sentence (a4) Color information by citing motivation type (B) Literature information search bar (C) Paper information view (D) Citing sentence topic and citing motivation view (E) Sentence details view. (Updated from [17]) (Demo URL: <https://paper-detector-9a67e.web.app/>)

the study of Dong et al. [5], and the words were clustered into ten groups (quantity, frequency, tense, example, suggest, hedge, idea, basis, compare, and result). However, as few cue words were suggested in this study and additional words required for this study were found, it was necessary to construct a cue word dictionary in addition to the suggested words. Therefore, this study searched for synonyms in WordNet [12] to collect additional words found during the study and to construct a dictionary for the added words. In addition, we have identified the syntactic features of general citations and used them as additional criteria when classifying the citing motivation. Generally, when describing the background of the research, it is usually written in an active sentence. And the basic methodologies and tools are mainly introduced in a passive sentence. Therefore, we found the main verb in the dependency of the sentence obtained through syntax analysis and checked whether the verb is active or passive to distinguish the voice of the sentences.

## 4 DATA VISUALIZATION AND INTERACTION

### 4.1 Visualization design task and interface

The interface proposed in this study should be able to visually summarize the various citation information data. Moreover, a figure overview is required to identify what citation relationships and patterns the InfoVis papers form. For this purpose, the design tasks defined in this study are shown in Table 2.

The visualization interface of this study consists of four views, the descriptions of which are as follows. The four views are driven by setting a specific target paper.

- Citation showing the full figure of the citation pattern and motivation Network overview (Figure 1 (A)).
- View showing the selected target paper (Figure 1 (C)).
- Table view showing the citation sentence topic and citing motivation (Figure 1 (D)).

Table 2: Task and design objectives for creating interface.

Task	Design object
1. Citation relationship and citation pattern identification	To confirm the overall citation relationships and patterns, the papers are listed according to the year of publication, and the target papers for further exploration are selected by providing interaction to select specific papers as the targets.
2. Citation sentence with motivation and topic summary identification	Details such as citation sentence, motivation, and sentence topic are provided to identify the meaning of the papers forming the citation relationship.

- Sentence card view showing the actual citation sentence and subtitle of the section in which the sentence appears in the paper (Figure 1 (E)).

### 4.2 Visualization interaction

In Figure 1 (a2), a bar graph depicts the number of citing papers by year. Furthermore, Figure 1 (a3) shows the citation network overview of the paper represented by square nodes. Papers with lower citations are arranged toward the right on the x-axis, while more recent years are shown higher on the y-axis.

Thus, the paper nodes are arranged in descending order of the number of citations in the row representing the year of publication. Figure 1 (C), provides a view showing summary information on the target (or 'seed') paper. If the user clicks on a specific paper node, that paper becomes the target paper and the metadata of the target paper can be viewed as shown in Figure 1 (a3).

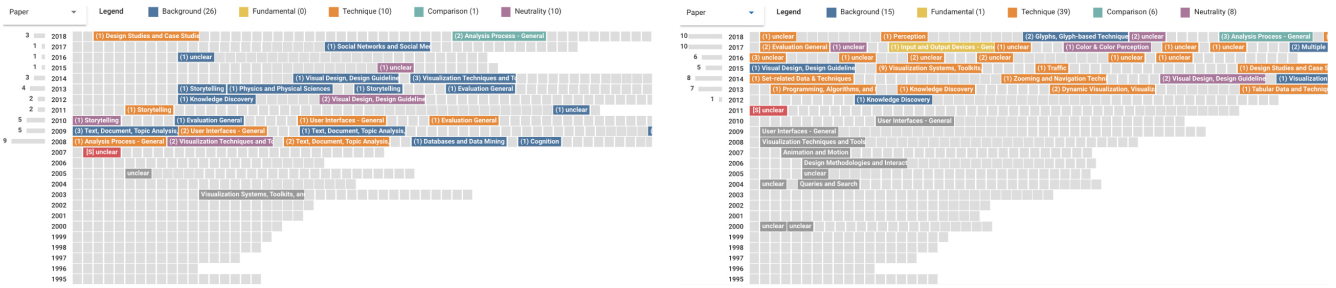


Figure 2: Paper base mode: (left) Overview with ManyEyes selected as the target paper; (right) Overview with D3 selected as the target paper.

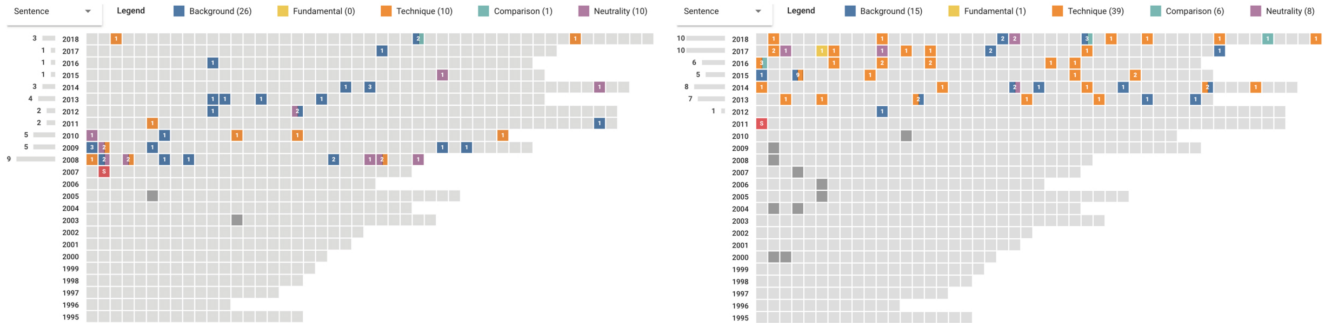


Figure 3: Sentence base mode: (left) Overview with ManyEyes selected as the target paper; (right) Overview with D3 selected as the target paper.

In the overview, the selected target paper is indicated in red and the paper nodes citing the target paper are shown in different colors. The colors are linked to the five citing motivations defined above, as shown in Figure 1 (a4). Papers referenced in the target paper are expressed in dark gray to indicate the relationship between citation and reference. If the paper that the user wishes to check is clearly present and the user knows its title, the title search function can be used (Figure 1 (B)). When searching, the position of the selected paper can be immediately seen in the overview.

The overview can be confirmed by selecting two options. First, the sentence-based option (Figure 1 (a3)) displays the square color node related with citation motivation. And each node presents the number of sentences of seed node mentioned. This makes it possible to infer the degree of association between the citing paper and the cited paper through the number of mentions in the citing paper. When a target paper is mentioned numerous times in one citing paper and each citing sentence is analyzed as having various motivations, the node displays the colors of all analyzed citing motivations. Also, the paper referenced by the target paper is indicated in dark gray.

Next, The paper-based option is to set the node attributes based on the paper research topic (Figure 2). When this option is selected, the activated node is a paper that cites the target paper and has a color corresponding to the citing motivation. Each color node includes the topic information. If the users want to see the full name of topic, they can confirm that at the paper information view (Figure 1 (C)). The paper-based option presents the research topics of the target and cited papers, thus the users are able to confirm the paper's theme affected by the target paper.

The user can also confirm the research topics of the papers that the target paper references and use this to examine which research topics are linked based on the target paper. The user can select the target paper to show a table view at the bottom of the screen. First column in the table indicates the topic of the cited sentence, and second column shows square nodes which indicates each cited sentences (Figure 1 (D)). The color of the node links the citing motivation analyzed in the citation sentence.

In the citing paper, the detailed information of the citation sentence referring to the target paper is provided in the form of a list as shown in Figure 1 (E). The list shows the group information of cue words by motivation, enabling the citing motivations to be interpreted in more detail. In addition, the research topic information extracted from the keywords of the sentence is provided so that the topic in the sentence can be identified without directly reading the sentence. Furthermore, a filter function is provided for the user to select and remove citing motivations so that they can examine only sentences with the desired citing motivation.

## 5 CASE STUDIES(VISUALIZATION IMPLEMENTATION)

Here, a case study is conducted to show the exploration process using the exploring system proposed in this study. The case study consists of two tasks and explains the usefulness of the results of this study.

### 5.1 Task 1. Can the citation relationships and citation patterns be identified?

Figure 2 shows how the paper-based option settings are set in the citation network overview. In this view, the representative citing motivation can be identified through the research topics and colors of the papers citing the target paper. The research topic of ManyEyes [16] is 'unclear'; this does not mean that it cannot be identified but that it is classified as a more general visualization study. However, as shown, ManyEyes is cited in papers with various detailed research topics, including visualization techniques, systems, storytelling, and interface. D3 [1] is also cited in papers with various research topics. Furthermore, the papers referenced by the target paper are displayed in dark gray below the target paper node, showing the citing and cited relationship based on the target paper. D3 referenced studies on general visualization and user interface.

Figure 3 shows the view when the node attribute of the citation network overview is set based on the citing sentence. Here, the node indicates the number of times the target paper appears in that paper

