Kairion: a rhetorical approach to the visualization of sources

ANDREAS KARATSOLIS
Writing Program Director
Albany College of Pharmacy
CL 206A -106 New Scotland Avenue
Albany, New York 12208
USA

Abstract: - Electronic systems which claim to support academic reading and writing practices consider only the descriptive metadata of sources as fields of interest for writers. However, both the relative lack of success of these systems and research in citation practices reveals that the motivation of academics when they use sources in their writing is primarily rhetorical. However, no detailed model for the rhetorical use of citations which could lead to the development of an electronic system utilizing rhetorical metadata has ever been developed. This paper reports on research which led to the development of such a rhetorical model. In this model, the dimension of time as well as the categories of values, relevance and significance are identified based on a combination of quantitative and ethnographic methods. This model is then used to develop the prototype for a rhetorical electronic system to support academic citation practices, *kairion.* The implications of such a system modeled for the work of individual academics in different stages of their career as well for groups of researchers are also discussed.

Key-Words: - **citation analysis, rhetorical metadata, visualization, electronic systems, source synthesis.**

1 Background

In the last few years we have witnessed a radical transformation in the way the academic community negotiates and exchanges information: vast digital libraries have been developed with the support of generous grants, proceedings from the most recent conferences are published within days at the sites of professional organizations, and the majority of professional journals are made available in electronic format to their subscribers. Since most academics retrieve the sources they need for their work using electronic means, and they also almost exclusively use computers for the own writing [1], the transition from paper to electronic text should follow naturally. However, paper still seems to be the preferred medium for reading academic text not only because of the traditional appeal of books as familiar objects [2], but also because of its portability and flexibility in spatial layout, as well as its navigation and annotation capabilities [3].

Recognizing this opportunity, designers of electronic systems have attempted to develop both hardware and software which would support the practice of reading and writing from sources. The introduction of Tablet PCs into the market, as well as a number of annotation systems (Annotator, CiteSeer) and several bibliographic databases (EndNote, Refvis, BibRex) came with the promise of supporting academics access, read, take notes, store and retrieve of information from sources. However, these systems have not gained the popularity of word processors or search databases. The reason seems to be that in most of these systems texts are treated as static objects with clearly identifiable attributes, or **descriptive metadata**, which can be used to store, retrieve and make connections between them in presumably meaningful ways: authors' names, publication, date, subject keywords, quotes, and summaries are all static fields that are already embedded within the document and are even sometimes available for download through a library's database. Even one's notes about a source need to be modified and placed in the context of the project or the argument put forward in each case. It seems that the underlying assumptions about the nature of texts that the developers of electronic tools have would have to be examined more closely.

In fact, research in rhetorical studies has challenged the traditional understanding about texts as static repositories of knowledge and instead proposed seeing them as "virtual objects," objects that do not yet exist except in the mind [4, 5]. Geisler [5] argued that texts play multiple roles in the mind of writers including being a driving motive or a desirable outcome. When texts are mediated by electronic systems, the virtuality of texts becomes even more evident, so the opportunity to develop

systems which will make this state transparent is enormous. However, focusing only on descriptive metadata limits to a great extent the potential for seeing them as virtual objects, creating meaningful attributes and connections between them, and eventually providing a context for the use of these sources in writing. These virtual, rhetorical objects transmit information, or **rhetorical metadata**, which can be equally or more important than descriptive metadata when academics decide which sources to use and how in their work.

Scholars studying academic attribution have actually been arguing for years that we need to understand the rhetorical dimension of citations in order to develop a theory of citation [6, 7]. Cozzens [6] argued that the aspect most likely to influence citations is rhetorical convenience, which is comprised of a rhetorical aspect and an allocation of credit or reward. Danette Paul [8] examined citations in the field of Chaos theory to conclude that there is no scientific, objective value of what constitutes a fact, and that scientists respond equally both to the quality of contribution and the quality of the rhetorical effort. Despite these significant insights, a detailed rhetorical model of citation practices, which might provide an understanding of the type of metadata academics find useful when reading and writing from sources.

This paper has a dual purpose: first it reports on the results of an ethnographic study of academics using citations in order to develop a rhetorical citation model. Then, it presents an example of the way this model can be incorporated into an electronic environment which will support the practices of both old-timers and newcomers in the academia.

2 Developing Rhetorical Categories

In order to arrive at a rhetorical model of citations, sixteen academics from four disciplines (Computer Science, Chemical Engineering, Materials Science Engineering and Humanities and Social Science) from a large research institute in the Northeast were recruited. They all provided access to at least two of their recently published journal articles to be analyzed for citation patterns and also agreed to a one-hour interview. Each interview included two parts: a typical source use narrative and a discourse-based interview, as it was originally developed in [9]. The typical source use narratives asked participants to tell the story of their reading and writing process when engaged in academic research. These narratives were meant to yield information similar to the typical use case scenarios that designers use to develop electronic systems. Immediately following this interview, they were asked to explain the decisions they had made to at least 15 instances of citation from their own writing. The questions were orchestrated in such a way that participants would have to explain if using a different citation pattern would have made a difference for the particular instance of use.

All the interviews were then coded automatically for keywords (nouns) associated with either to descriptive or rhetorical metadata (Table 1).

Descriptive	Rhetorical
Note	Work (noun)
Title	Research
Author	Field
Abstract	Project
Keyword	Argument
Date	Reader
Topic	Group
Quote	People

Table 1: Autocode analysis terms

The descriptive terms can typically be identified as objective characteristics of a source, and several electronic systems have used them as fields. The rhetorical terms, on the other hand, are all related to the rhetorical context surrounding a source such as the audience or the argument presented.

The analysis of the two types of interviews showed significant differences in the way participants described their practices. In typical source use narratives, participants used more descriptive terms than rhetorical terms (Figure 1).

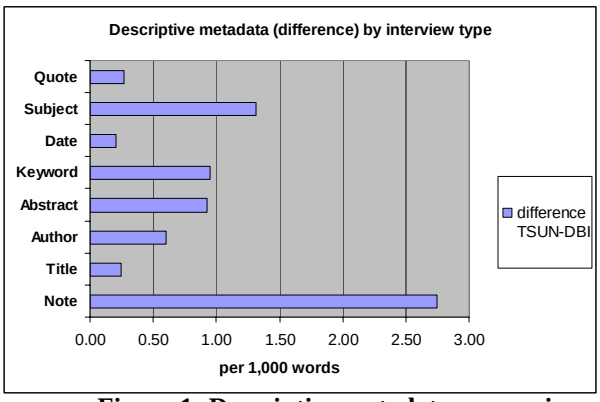

Figure 1: Descriptive metadata comparison

On the other hand, in discourse-based interviews, which asked more direct questions about instances of citation use, participants used more rhetorical terms to describe their citation practices (Figure 2). The terms "project" and "research" were the only ones which were more prevalent in typical source use narratives, but they were used to refer both to the participants' own work and the work they were citing, so further analysis was necessary.

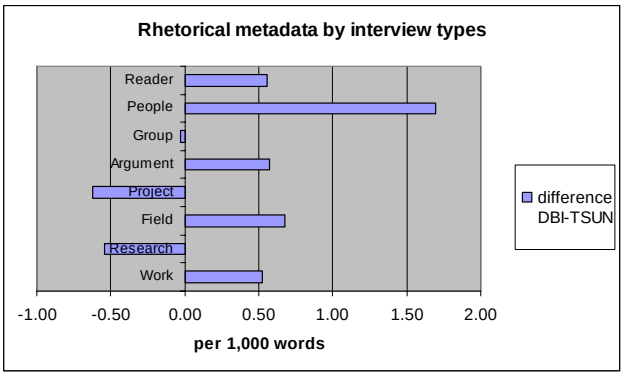

Figure 2: Rhetorical metadata comparison

A further analysis of the rate of use of these keywords showed that while overall for the typical source use narratives both categories of keywords were almost equally present in the words of the participants, the picture was very different for the discourse-based interviews. When participants were asked about specific instances of use, the language they used was overwhelmingly focused around rhetorical considerations with a ratio of 1:6 (Figure 3). These results clearly point to the direction that rhetorical considerations are the basic motivation behind the use of sources in most instances, at least for these sixteen participants. These results were both consistent with the literature and can explain why the current electronic systems aimed to support reading and writing have not been so popular.

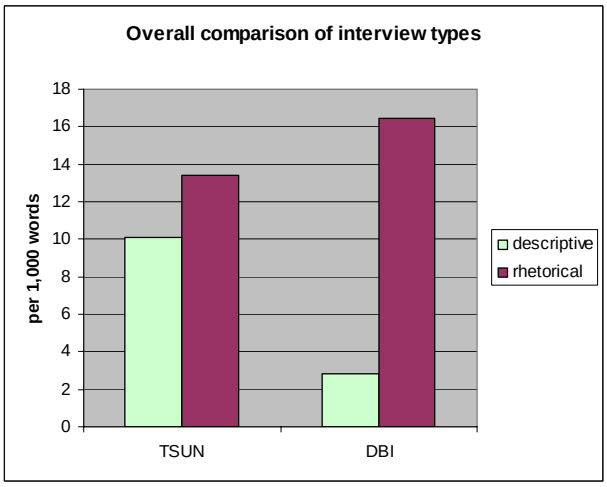

Figure 3: Comparison of metadata

The next step in the analysis was to identify the general categories that participants used in the discourse-based interviews to explain their citation decisions. After analyzing the responses to almost 250 questions, the first pattern that became obvious was that in most instances the decisions to cite a source were directly related to the project itself in relation to the active life of the research community. Readings and consequent writings are always embedded within a context of purpose and audience, as well as a sense of the appropriate moment (or *kairos* in ancient Greek rhetoric) for the field or discipline.

Beyond the dimension of appropriate time, participants identified a big number of critical elements which compelled them to cite sources in their work. As an example, one participant described a decision to include a source like this:

"I think in my field this is probably a fairly well-known statement by now, but you're almost referencing a source because of people like [my co-author] who look at the optimization side of things and aren't going to know this…"

Such a statement shows that a consideration for the audience's level of knowledge about the topic was very important, especially since this was interdisciplinary work. If this particular concern was repeated in several of the participants' responses, then I was able to formulate a general evaluative statement of the type: "This source is a good introduction to the problem for new readers/participants."

After combining all the discrete statements developed from the participants' responses, three general categories emerged, which corresponded well with the results of previous research:

- **Value** (contribution to the knowledge of the field)
- **Relevance** (topical relation to the current argument/project)
- **Significance** (role in shaping the problem for the audience /field)

Examples of statements from the category of value included:

- This source has been heavily cited in your field
- This is one of the first/original/pioneering/classic papers in your field

- This source makes an important contribution modification/improvement to the knowledge of the field.

Examples of statements from the category of value included:

- You have authored/co-authored or are affiliated with the authors/group of this source
- Your current project extends some of the work presented in this source
- Your current project attempts to overcome some of the limitations of the work presented in this source.

Examples of statement from the category of value included:

- This source is authored by an individual or group who is at the forefront of innovation/development of new ideas in the field
- This source was published in a well-respected journal or by a well-respected publisher in your field.
- This source is so central in the field/for the problem at the moment that your reviewers/readers will expect to see a discussion of it
- This source is very appropriate/will be well-received for the publication venue this document will be submitted to.

All these statements in the three categories could then be modeled into an electronic system which would provide not only the descriptive dimension of a source, but also its rhetorical one. In the next section I will present *Kairion*, which stands for "the appropriate, critical" either moment in time or rhetorical consideration of value, relevance or significance which determines when and how a source will be used.

3 Visualizing rhetorical metadata

The next step in this research was to show that these results, the categories and the associated questions could actually be turned into metadata which academics could use to support their reading and writing practices. For this reason, the original statements that were developed out of the participants' responses were turned into items to be rated by the readers. However, the data points for each source were more then twenty, making it very difficult to help the reader "see" all the rhetorical considerations. Therefore, the three general categories were utilized to show an aggregate of the ratings, instead of individual data points. This way the schematic of a hexagon would be developed if the source received the full rating for all the questions in all the categories, since there are three categories presented in the dimensions of present and past time.

In the following figure (Figure 4) the way the visualization was conceptualized is presented. Both the dimension of time, and the three categories can be shown as data points on the hexagon.

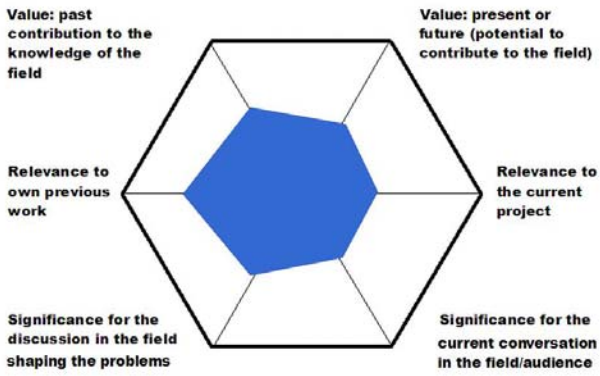

Figure 4: Visualization model of *Kairion*

This basic model was then developed into an application where the where the user, after inputting the descriptive metadata of the source, is asked to rate it (1-5) according to the questions associated with each category. There are only four questions for each dimension of the category, for a total of 24 questions. Figure 5 shows the ratings view of the application.

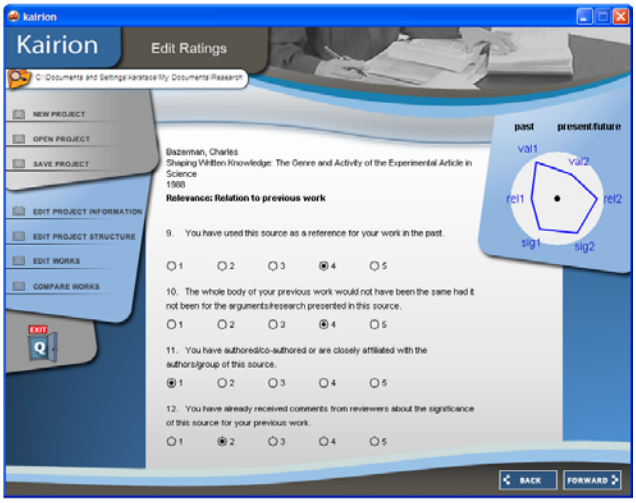

Figure 5: Ratings view in *Kairion*

Using Macromedia Director and Flash technology, the hexagon on the upper right-hand corner is

constructed as the user is rating the source. This visualization can then be used as a guideline as to what where and why a citation fits in the overall structure of the document or the specific argument the author is making. However, precisely because this information is difficult to recapture at a later time without going through all the visualizations of all the sources again, another view showing the works within the project structure is available (Figure 6). In this view, users can easily connect a certain type of source to their project's outline, for example connect all sources with a high past value (the "classic" sources in the field) to the background section of the document.

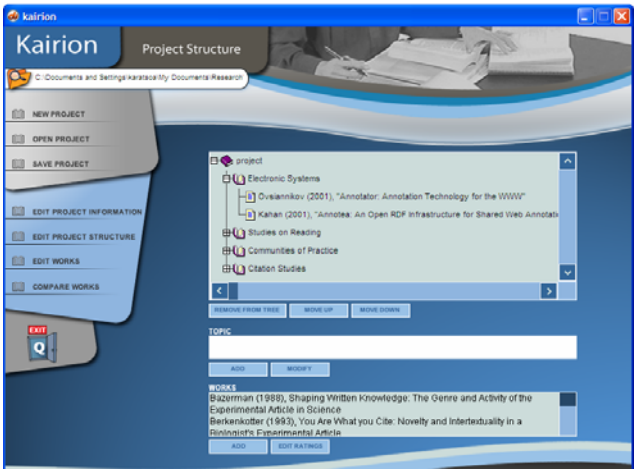

Figure 6: Project structure view in *Kairion*

Finally, because very often comparisons between sources can be very useful in determining their relative importance for one of the categories, another view where two visualizations appear either separated or overlapping is available. This view can also show the differences between two readings of the same source, perhaps from an advisor/student pair, or by the same individual for two different projects.

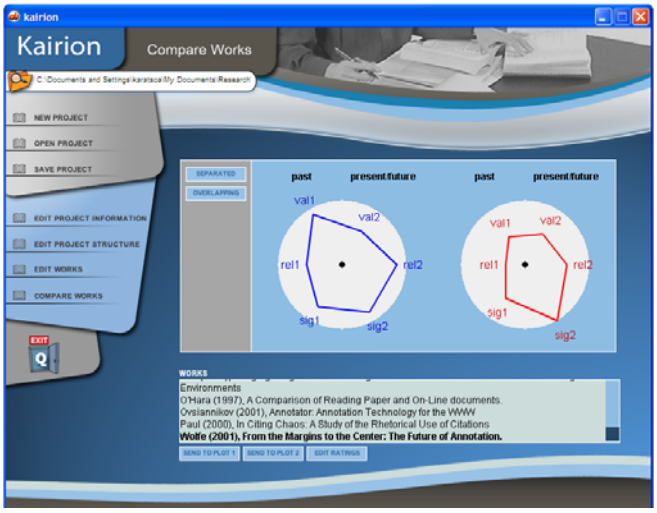

Figure 7: Comparative view in *Kairion*

For the purposes of this discussion, these are the most significant views of *kairos*, which show how it can support rhetorical metadata. However, since it utilizes an XML backend, the possibilities of repurposing the information into other visualizations are open.

4 Conclusion

Kairion is only an example of an electronic system which attempts to visualize the rhetorical dimension of sources so that academic practitioners can make more effective decisions about the sources they will use and the way they will use them. Still in prototype phase, K*airion* would have to be introduced to users and its use has to be analyzed for the duration of a project in order to arrive at conclusions about the accuracy of the three categories and the dimension of time. An item analysis of the individual questions will also have to be performed, ideally with users from different disciplines and stages in their academic career.

The implications of such a system for individual practices are obvious, as academics very often reuse sources for multiple projects or multiple documents coming out of one project. However, for groups of researchers, a common database of rhetorical metadata where the "readings" of the different group members would be transparent has important advantages.

Finally, *Kairion* and similar systems can serve as learning environments where peripheral participants in a field will be able to see in a transparent way how full participants understand the role of a source for their work and for the field. This way, such electronic systems can serve not simply as repositories of information, but as flexible objects aimed at supporting real practices.

References:

[1] Eileen Wood, Tina Willoughby, Lisa Porter and Jacqueline Specht, An examination of how a cross-section of academics use computer technology when writing academic papers, *Computers in Education*, Vol.38, No.1-3, 2002, pp. 287-301.

[2] David M. Levy, *Scrolling Forward*, Arcade, 2001

[3]Abigail Sellen and Richard Harper, *The Myth of the Paperless Office*, MIT Press, 2003

[4] Peter Medway, Virtual and material buildings: Construction and constructivism in architecture and writing. *Written Communicatio*n, 13(4): 1996, pp. 473-514.

[5] Cheryl Geisler, (2001). Textual objects: Accounting for the role of texts in the everyday life of complex organizations. *Written Communication ,18*, 2001, 296-325.

[6] Susan Cozzens. What do Citations Count? The Rhetoric-first model, *Scientometrics.* Vol.15, No.5-6, 1989, pp. 437-447.

[7] Blaise Cronin, *The Citation Process*, Taylor Graham, 1984.

[8] Danette Paul, In Citing Chaos: A Study of the Rhetorical Use of Citations, *Journal of Business and Technical Communication,* Vol. 14, No.2, (2000), 185-222.

[9] Lee. Odell, D. Goswami, and A. Herrington. The Discourse-based Interview: A Procedure for Exploring the Tacit Knowledge of Writers in Nonacademic Settings. In P. Mosenthal, L. Tamor, & S. Walmsley (Eds.). *Research on Writing: Principles and Methods*, Longman 1983, 221-235