BIVTECI: A Bibliographic Visualization Tool

David Modjeska, Vassilios Tzerpos, Petros Faloutsos, Michalis Faloutsos

August 5, 1996

Contents

Abstract

Relatively little research has been published about bibliographic visualization, although scientists, researchers, and students often have need for such technology. In this paper, we examine the functionality needed to visualize bibliographies effectively. After reviewing some existing tools, we present BIVTECI (Bibliographic Visualization Tool with Enhanced Citation Interactivity) which derives from our analysis.

Our main research goal is to capture the user-required relationships between bibliographic entries or articles. First, a user must understand the chronology of articles. Second, a user needs to see the influence of articles on later work, i.e., the citation relationships. Third, the user wishes to group articles by specified attributes (e.g., title, author, keyword). Fourth, the user needs to examine information at varying levels of detail from general, multi-article views to specific, single-article ones. Finally, the user may need to use several information views simultaneously. We note that baseline functionality includes simple retrieval of entries by title, author, and keyword.

The prototype BIVTECI tool demonstrates effective solutions for most of these requirements, combining information visualization with manipulation. The article proposes future research into features

not yet implemented. BIVTECI is based on the Star data translator and the Landscape information visualizer, both developed at the University of Toronto.

1 Introduction

An essential tool of scientific research, bibliographic cross-referencing and searching are often extremely time-consuming. From a particular article, researchers must follow reference links backwards to cited articles and forwards to citing articles, in order to discover relevant methods, conclusions, and further references. Unfortunately, general-use tools tend to be slow. Paper and electronic catalogues, for example, support the user in following links. Library catalogues support backward linking, while citation indices support forward linking.

Existing general bibliographic tools have only minimal support for interactivity and visualization. Catalogues tend to list detailed entries, or at best to filter entries according to specified attributes. Researchers, scientists, and students in many disciplines could benefit from more sophisticated bibliographic tools. Numerous situations require such tools:

- research in an area that is partially or entirely unknown
- survey or overview research
- monitoring of research activities in specific fields over time
- examination of relationships among articles in a specific field
- maintenance, visualization, and searching of a personal database

A good bibliographic visualization tool (BVT) must present the key relationships between bibliographic entries or articles. At a minimum, we propose that such a tool support the following six features:

1. display of complete bibliographical entry
2. filtering by title, author, and keyword
3. display of chronology and influence of articles, i.e., ordered citation links
4. information views at several levels of detail
5. multiple simultaneous and (optionally) synchronized views
6. visualization of large search result sets

In addition, a BVT could effectively support the following three features:

1. article clustering
2. attribute visualization
3. system inference of articles related to user queries

Bibliographic data is characterized by strongly networked relationships. For this reason, a standard set-oriented or relational database cannot adequately visualize or manipulate such data. Moreover, a generalized BVT could serve as a front end for any networked data set, e.g., a local database, or World Wide Web documents or servers. A BVT, therefore, could serve as a springboard for future research in the area of visualization.

In the Department of Computer Science at the University of Toronto, we have therefore developed the software prototype BIVTECI – Bibliographic Visualization Tool with Enhanced Citation Interactivity. The prototype serves three purposes. First, it implements a growing subset of the features proposed for an effective BVT. Second, it presents a general user interface (UI) for strongly networked data sets. Finally, it demonstrates the effectiveness of the Star data translator and the Landscape information visualizer (which will be discussed in Section 4 of this article).

BIVTECI supports three bibliographical information views. The general view is a chronological overview of articles with citation links. The specific view centers on a specific article, while displaying its cited and citing links. Finally, an intermediate view arranges articles spatially by attribute relevance. Several features are shared among all views. First, each view supports filtering by title, author, keyword, and other attributes. Each article displays, in iconic form, details of its participation in the current query. Second, each view supports article aggregates on the basis of various attributes. Third, each view supports *virtual references*, or system-inferred links related to (but not resulting from) the user's current query. In general, views can be independent or synchronized; that is, filters can have local or global scope. As required, full bibliographic information is immediately available for each displayed document.

To accomodate a range of bibliographic expertise from novice to expert, BIVTECI features several UI and retrieval options. Such options include attribute weights, attribute displays, retrieval parameters, and view layouts. The system provides useful default values for the novice user.

The remainder of this paper is organized as follows: Section 2 presents an overview of previous work on the visualization of bibliographic data; Section 3 discusses in detail the information views supported by BIVTECI; Section 4 describes the system's architecture; and Section 5 summarizes our research and proposes future work.

2 Review of Bibliographic Visualization Tools

Researchers at Xerox PARC (Palo Alto Research Center) have developed two tools useful for bibliographical work. Both tools have been widely publicized, so this article will discuss them only briefly. (For references, please see Section 6.)

For the last several years, PARC researchers have collaborated on the Information Visualizer, an application based on 3D graphics and interactive animation. Part of this work, the Butterfly system is an application for accessing DIALOG's Science Citation databases over the Internet. The application integrates search, browsing, and access management. Sophisticated in visualization and access techniques, Butterfly meets the requirements proposed above for a BVT. Utilizing 2D graphics and standalone workstations, BIVTECI meets the requirements for a BVT somewhat differently, while incorporating the three additional features recommended above. [Mackinlay 95]

Another project at PARC, TileBars, introduces a new visualization technique in the field of information retrieval (IR). The technique uses text structure to visualize retrieval from full-text documents. TileBars simultaneously and compactly display key attributes of retrieved documents. The visual patterns can be quickly scanned and interpreted, helping users to judge the potential relevance of documents. Working with bibliographic entries, BIVTECI adapts the TileBars technique to indicate query attributes (e.g., title, author, keywords) pertaining to displayed articles and clusters. [Hearst 95]

Several researchers at the University of Padua and at the European Space Agency have designed and implemented a two-level conceptual architecture for constructing a hypertext to interact with large textual databases. The architecture is intended to be a semantic representation of both the content and the organization of a document collection. A general goal is to define a new model for information retrieval (IR). The resulting prototype, HYPERLINE, integrates a thesaurus with bibliographic retrieval. While it also enhances bibliographic retrieval with browsing functionality, BIVTECI organizes browsing by citation links, rather than by semantic ones. In addition, BIVTECI incorporates graphical visualization techniques. [Agosti 92]

At GMD-IPSI in Darmstadt, researchers have developed a prototype UI named LyberWorld. The prototype visualizes an abstract information space for the full-text IR system INQUERY. To assist user searching and browsing, LyberWorld's visualizations use metaphors of 3-D spatial navigation. One of these visualizations, the RelevanceSphere, supports relevance feedback and document clustering. BIVTECI's intermediate, relevance view resembles aspects of a hypothetical 2-D RelevanceSphere.

3 Views

3.1 Overview

As proposed, our prototype BVT implements three information views. Each view presents bibliographic data at a different level of generality, and each emphasizes different relationships and attributes. We believe that the current views compose an adequate foundation for many potential applications. The system can be easily extended, however, to support new views. (see Section 5).

As previously mentioned, the system generates views on the basis of attributes. Such attributes include title, author(s), year, location, journal/proceedings, publisher, subject/area, and keywords. Of these attributes, keywords have particular value for categorizing bibliographic elements, and for indicating associated subjects and research areas. Keywords are normally supplied by authors and are often included in bibliographic databases. Consider, for example, the keywords used in the bibliography files (*.bib) supported by BibTeX, a bibliographical enhancement for LaTeX, which is a popular and powerful text formatting tool. BIVTECI can process BibTeXfiles to identify the keywords. We are currently investigating automated ways to generate keywords and to identify subjects on the basis of other available information. At the moment, relationships between bibliograhpic entries and subjects/areas are determined primarily by keywords. For each entry, the system has a list of weighted keywords. Weights are currently appended by a system user during data entry. In the future, however, we hope to consider IR techniques to initialize weights automatically by full-text analysis.

The items displayed by BIVTECI are labeled (by default) with the first two letters of the authors last name, and the year of publication. We follow this convention in the next sections when discussing a sample bibliographic database. For the sake of visual clarity, this articles database contains only fourteen papers.

3.2 General View

A user often needs to visualize an entire bibliographic database or a large subset according to specific criteria. Such a visualization should show the general chronology, dependencies, and attributes of bibliographic entries. This requirement has led to the development of a general overview. The user initially specifies such attributes as subject, author, etc.; the system then presents matching bibliographic items in chronological order.

Figure 1 shows a visualization of a complete, small database. The top row contains the most recent papers (in this example from 1995), and the arrows indicate reference relationships with arrows pointing to the cited papers. The user can customize the view in various ways. For example, he or she can constrain the system to organize items into distinct rows by publication year (as shown in Figure 1). The user can also choose to display (as shown) or supress citation links. Links can also be filtered by time interval, e.g., the system could display only links spanning two or fewer years. In addition to filtering items by attribute, the system can also highlight items. In Figure 1, for example, the gray boxes highlight papers by a specific author. Using color (or grey scale), BIVTECI can also represent

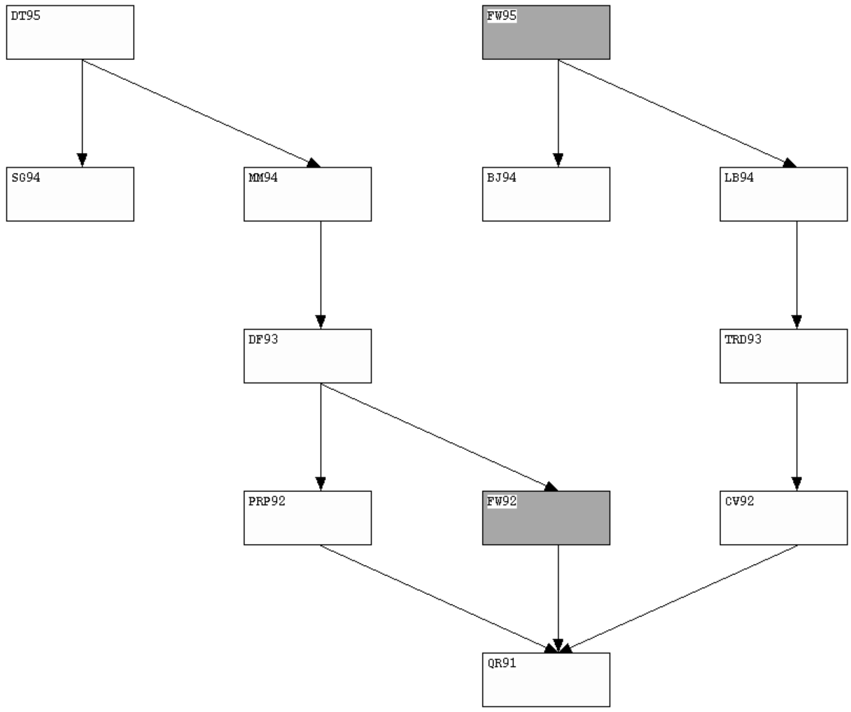


Figure 1: General view. Items with reference links, sorted chronologically. Highlighting indicates a specific author.

papers with multiple attributes. An item possessing the requested attributes displays an appropriately striped band near one side. Figure 2 shows an example of a bibliographic entry with attribute bars, as well as an on-screen legend for the attributes.

In the future, the system will also use colors to indicate papers within user-specified ranges of attribute values. For example, a paper's impact might be estimated by the numbers of times it is referenced. Articles cited more than ten times might be displayed as red-shaded rectangles, while articles cited only once might be displayed with blue shading.

BIVTECI supports a feature that is called *virtual references*. Virtual references are citation links that do not exist, but they are possibly implied by subject and keywords. Such reference hints from the system can indicate unreferenced but relevant previous research. In contrast to the bold lines of regular citation links, virtual references are displayed using light or dotted lines. (Please see Figure 3.) We plan to enhance the functionality of virtual references in the future.

At any time, the user can obtain a full bibliographic description of a displayed item by clicking on it with the mouse. The system retrieves relevant information from the database and displays it in a pop-up text window. In a similar way, during overview browsing, users often wish to see detailed citation information about a specific paper. To accommodate this need, the system allows the user to select a paper and to invoke a detailed view, which is described in the next section.

3.3 Specific View

BIVTECI's specific view presents the relations between a bibliographic item and its citers or citees. This view supports detailed bibliographic research. Figure 3 (a) presents an example of such a view for the article labeled "GM93". The articles that reference (or could reference) GM93 are shown at the

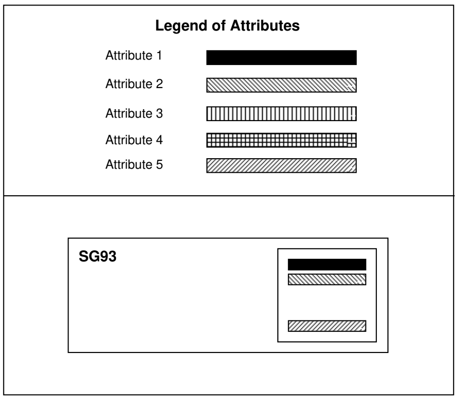
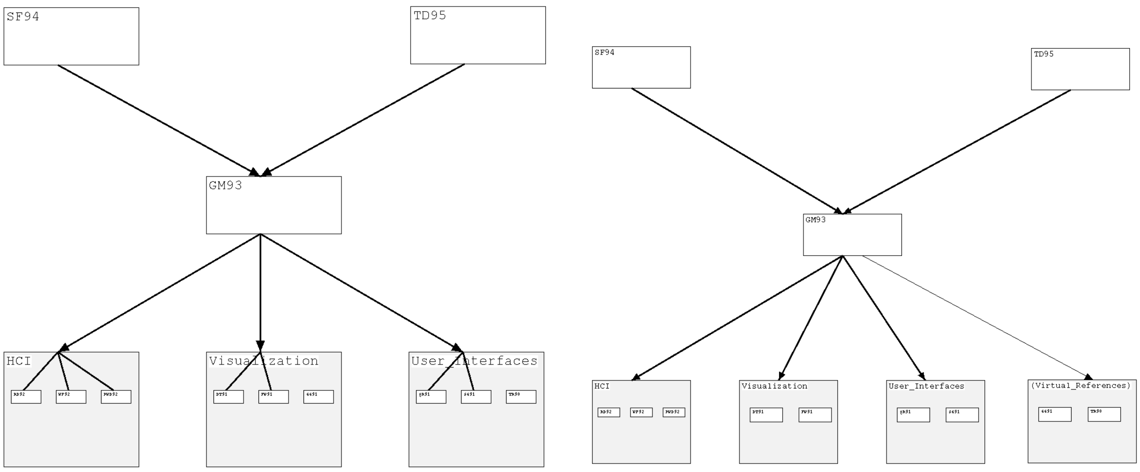


Figure 2: Attribute bars. On-screen legend and sample item.



(a) (b)

Figure 3: Specific view. Items with citing articles, cited groups, and virtual references.

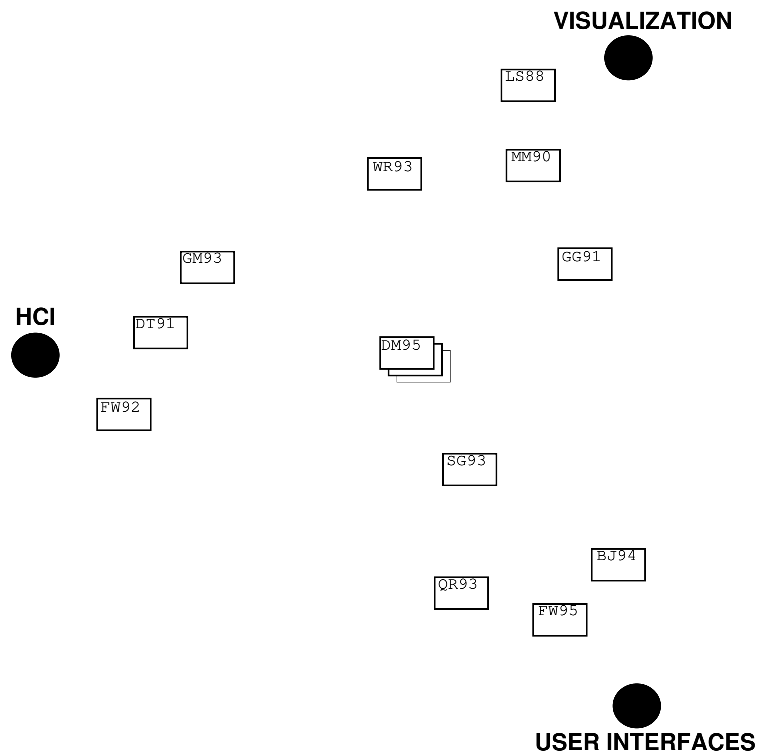


Figure 4: Relevance view. Items arranged by keyword relevance to three topics.

top, and the articles that are (or could be) referenced by it are at the bottom. The top row is sorted by attribute, while the bottom consists of articles grouped by topic. These article groups support both task-level thinking and compact information display. In the example, three areas are shown – *Visualization, Interfaces* and *HCI* (Human-Computer Interaction). Virtual references can be displayed as individual items in other groups (as shown by the unlinked group members in Figure 3 (a)) or as a distinct group (as in Figure 3 (b)).

As in other views, the user can specify attributes to filter, highlight or color items. For example, the view can be restricted to contain the papers written by a certain group of authors during a specific time interval.

3.4 Relevance View

The previous two views are organized by bibliographic entry, with attribute information used primarily for filtering and highlighting. We found it useful to develop a view organized by attributes, keywords in particular. Such a view should support reasoning by relative attribute weights. For this purpose, BIVTECI provides an intermediate, relevance view that presents information in a non-hierarchical, multi-focus layout. The user specifies one, two or three attributes which form a point, a line or a triangle, respectively, in the plane. Each bibliographic item is then placed on this plane according to its relationship to the given attributes. Figure 4 shows an example of such a view where three foci are used to spatially organize the database. Proximity to a focus indicates article relevance. To facilitate the spreading of articles, the user can adjust the relative weights of the foci. While providing useful relevance feedback, the layout of the relevance view is approximate. As an aside, as shown in Figure 4, a stack labelled by the top article indicates that several articles are co-located at this point of the plane.

By clicking on a stack, the user can examine the items it contains.

As in other views, filtering, coloring and highlighting can be employed by the user to restrict or reveal information about the displayed articles. Similarly, the user can choose to show or hide citation arrows between bibliographic items.

4 System Architecture

BIVTECI was developed at the University of Toronto using tools of the Systems group in the Department of Computer Science. These tools have been used with much success, primarily for research in software engineering.

For its back end, BIVTECI uses the Star data translator. The tool translates between a variety of formats, including PROLOG facts, C source code, and Turing source code. In addition, Star can perform useful relational algebra operations. These operations support the layout features of BIVTECI.

For its front end, BIVTECI uses the Landscape information viewer. Landscape reads a file in a dedicated format, from which it generates a drawing. The user can then modify the diagram manually in many ways. Landscape can also be used to produce a new diagram, which is input to Star to generate file templates in Object-Oriented Turing (OOT) or C. Other features of Landscape include manual clustering of nodes, diving into a cluster, and first-order visual queries.

5 Conclusion

An essential tool of scientific research, bibliographic cross-referencing and searching are often time-consuming. Moreover, in existing general-use bibliographic tools, interactivity and visualization are only minimally supported. Knowledge workers in many disciplines could benefit from more sophisticated bibliographic tools. Numerous situations require such tools. A good BVT must present the key relationships between articles. At a minimum, we have proposed that such a tool support the following five features: display of bibliographical entries; data filtering; display of chronology and citations; specific and general views; simultaneous, synchronized views; and visualization of large result sets. In addition, a BVT could support the following three features: article clustering; attribute visualization; and system inference of query-related articles.

Bibliographic data is characterized by networked relationships. For this reason, a standard database cannot adequately visualize or manipulate such data. Furthermore, a generalized BVT can serve as a front end for any networked data set. We have therefore developed the software prototype BIVTECI. The tool implements a subset of an effective BVT; it presents a UI for networked data sets; and it demonstrates the effectiveness of the Star and the Landscape tools.

BIVTECI supports a chronological, multi-article overview; a single- article view; and a spatial attribute-relevance view. Each view supports filtering; each displayed article can indicate bibliographic attributes; and each view supports virtual references. In general, information views can be independent or synchronized, and complete bibliographic entries can be easily displayed.

From the perspective of HCI, current results confirm the usefulness of BIVTECI. As proposed, the required features have allowed for flexible and effective bibliographic research in small-scale trials. The systems graphical user interface (GUI) supports interactive exploration of citations and attributes. Furthermore, the three network visualizations allow a user to work in the high-level task domain of articles and citations, rather than at the low level of bibliographic entries.

From a systems point of view, the architecture of BIVTECI has proven to be successful. The two-component design allows for power and extensibility. The Star translator insulates the visualizer from dependence on specific data formats, thereby increasing the potential applications of BIVTECI. Moreover, the Landscape visualizer has enough generality to accommodate a range of information views, which allows for both incremental refinement and potential innovation.

Future work on BIVTECI encompasses three tasks. First, further development work is required to implement the full set of BVT features. Second, usability testing is needed to validate and refine the system's UI. Finally, performance testing would be useful to investigate the characteristics of the system.

The ultimate extension of BIVTECI, of course, would be real-time, on-line, full-text retrieval. Such research is beyond the scope of this paper. This BVT tool, however, can be seen as part of the growing research on digital libraries. We hope, therefore, that BIVTECI plays a useful role in this large and important project.

6 References

[Agosti 92] Agosti M., Gradenigo G., and Marchette P.G. A Hypertext Environment for Interacting with Large Textual Databases. Information Processing and Management. 29 (3): 371-387.

[Hearst 95] Hearst M.A. TileBars: Visualization of Term Distribution Information in Full Text Information Access. In Human Factors in Computing Systems: CHI 95 Conference Proceedings. New York: ACM, 1995.

[Hemmje 94] Hemmje M., Kunkel C., and Willett A. LyberWorld - A Visualization User Interface Supporting Fulltext Retrieval. In SIGIR '94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval. London: Springer-Verlag, 1994.

[Mackinlay 95] Mackinlay J.D., Rao R., and Card S.K. An Organic User Interface For Searching Citation Links. In Human Factors in Computing Systems: CHI 95 Conference Proceedings. New York: ACM, 1995.

[Mancoridis 94] Mancoridis S., Holt R.C., and Godfrey M.W. A Program Understanding Environment Based on the "Star" Approach to Tool Integration. In ACM CSC. New York: ACM, 1994.

[Penny 92] Penny D.A. The Software Landscape: A Visual Formalism for Programming-in-the-Large. Ph.D. Thesis, Department of Computer Science, University of Toronto, 1992.

[Shneiderman 92] Shneiderman B. Designing the User Interface: Strategies for Effective Human-Computer Interaction. New York: Addison- Wesley, 1992.

7 About the Authors

David Modjeska is a Ph.D. student in the Department of Computer Science at the University of Toronto. He works in the area of human-computer interaction, with particular interests in information visualization, hypermedia, and cyberspace. After completing a B.A. in English at Harvard University in 1982, Modjeska earned an M.S. in computer science at Stanford University in 1994. He also worked for six years as a software engineer in Silicon Valley, mostly developing GUI front-end tools for a relational database system. His Internet address is modjeska@dgp.toronto.edu.

Vassilios Tzerpos is a Ph.D. student in the Department of Computer Science at the University of Toronto. His research interests include software architecture, reverse engineering and software maintenance. Tzerpos received his B.Sc. from the Electrical and Computer Engineering Department at the National Technical University of Athens, Greece, in 1992 and his M.Sc. in computer science from the Department of Computer Science at the University of Toronto in 1995. His Internet address is vtzer@cs.toronto.edu.

Petros Faloutsos is a Ph.D. candidate in the Department of Computer Science at the University of Toronto, where he is currently working on computer animation, control and dynamic modeling. He received his B.Sc. in electrical engineering from the National Technical University of Athens, Greece, in 1993; he received his M.Sc. in computer science from the University of Toronto in 1995. His Internet address is pfal@dgp.toronto.edu.

Michalis Faloutsos completed a five-year B.Sc. in Electrical and Computer Engineering at the National Technical University of Athens, Greece, in 1993. He obtained an M.Sc. in Computer Science at the University of Toronto in 1995. He is currently studying in the Ph.D. program of the Department of Computer Science at the University of Toronto. Michalis' interests include computer networks, multicasting and volleyball. His Internet address is mfalou@cs.toronto.edu.