Scandinavian Journal of Information Systems

Volume 5 | Issue 1 Article 2

1993

VISUALIZATION OF A DOCUMENT COLLECTION WITH IMPLICIT AND EXPLICIT LINKS: The Vibe System

Kai A. Olsen
Molde College, Norway, KaiAOlsen@emailaddressnotknown

Robert R. Korfhage
University of Pittsburgh, RobertRKorfhage@emailaddressnotknown

Kenneth M. Sochats
University of Pittsburgh, KennethMSochats@emailaddressnotknown

Michael B. Spring
University of Pittsburgh, MichaelBSpring@emailaddressnotknown

James G. Williams
University of Pittsburgh, JamesGWilliams@emailaddressnotknown

Scandinavian Journal of Information Systems, Vol. 5, pp. 79–95, 1993

VISUALIZATION OF A DOCUMENT COLLECTION WITH IMPLICIT AND EXPLICIT LINKS

The Vibe System

Kai A. Olsen
Department of Computing Science, Molde College
N-6400 Molde, Norway

Robert R. Korfhage, Kenneth M. Sochats
Michael B. Spring, James G. Williams
School of Library and Information Science, University of Pittsburgh
Pittsburgh, PA15260, U.S.A.

Abstract

We study a document collection based on the implicit associations among the documents and on the explicit links, such as hypertext links, citations or author names, that exist among the documents. The paper presents a visualization method that gives an overview of a document collection, where both implicit and explicit links are shown. Both hypertext and non-hypertext document collections may be visualized. The technique presented shows promise in providing a tool for presenting a holistic view of a document collection.

Keywords: visualization, document retrieval, hypertext.

1 Introduction

This paper looks at collections of documents in order to *visualize* the different types of associations that might exist among the documents. The idea is to give a user visual clues that can be used for studying and organizing the documents in a collection, i.e., to present a user-defined structure of the collection. Although our focus in this paper is inter-document associations, the principles outlined could easily be applied for intra-document visualization. The tool described may also be applicable to visualizing existing hypertext documents in such a way as to assist in the authoring and browsing process, and to address the problem of becoming "lost in hyperspace" (Conklin 1987).

There are implicit and explicit links in any document corpus. The task of creating implicit links is similar to that of determining the results of a *query* in traditional information retrieval, where the task is to determine relevance between a document and a query (van Rijsbergen 1979). In fact, a query may be seen as a method of linking documents. Traditional methods include the use of content similarities, often defined by term to term relationships. These relationships may be determined either directly or indirectly through term clustering or by the use of a thesaurus.

The *explicit* links within a document corpus are links created by the author or derived from these. The types of explicit links that one may expect to find in a document will depend on the document category. For example, in a scientific journal paper we may find explicit links of the following type:

- Citations and other document references.
- References to data sources
- Author, organization, funding-support links. This may include primary and secondary authors, multiple support sources, etc.
- Links to the parent publication—journal, book, etc.
- Index terms, glossary, thesaurus

The implicit links in a document corpus exist on the basis of term to term relationships. They may—in principle—be converted *automatically* into explicit links. An interesting comment on this process, however, is made by Glushko (1989):

> "When we first began working in hypertext several years ago, we expected that it would soon be possible to extract these implicit links automatically with natural language processing or clever indexing techniques ..., but we have been disappointed so far and we are starting to conclude that implicit intra-document links are best identified by the hypertext reader."

We would argue that this comment is valid for implicit inter-document links as well. Implicit links would have to be created automatically on the basis of similarity on a lexical level (string matching). There can only be a *probability* that such lexically derived links will be valid at a semantic level. As Glushko concluded, it is important that implicit links reflect the needs of the user. It therefore seems reasonable to let the reader identify these links. Glushko uses an analogy to a used textbook to explain this point, "... the highlighting and margin notes [in a used textbook] may be useful to another student, but may be distracting or misleading at other times." This suggests that implicit links may be best thought of as a temporary and approximate link structure reflecting the dynamic interests of the user.

Thus we see that implicit and explicit inter-document links have different characteristics. We present a system based on these differences and offering a new perspective based on the different characteristics of implicit and explicit inter-document links, a system is presented that uses these differences to offer a new perspective for viewing document collections. The implicit links are presented by *positioning* documents in a user-defined information space. Explicit links, as references between documents, may then be presented as an overlay to this perspective.

2 Visualization

Visualization has been used to advantage for presenting the explicit link structure in hypertext documents (Conklin 1987). In such *maps*, a simple line will intuitively present a link between two nodes, and quite complicated structures can be presented in a limited space. Thus, visualization may give the reader an overview over a collection of document nodes that would be impossible or extremely difficult to get from text alone.

If implicit links could be visualized as well, in a user-oriented space, we would get a holistic visual presentation of the document or node collection. The idea would be to present the user with an overview of a document collection, giving as much data on the documents as possible.

Several different methods have been used for visualizing document collections, such as using projections from high-dimensional spaces, rotation and rendering techniques (Doi *et al.* 1991, Young & Rheingans1991). The SemNet system (Fairchild *et al.* 1988) allows for user placement of data points and the use of a centroid heuristic in placement of other points. However, there are disadvantages to these methods. Visualization of multi-dimensional spaces, such as those used for modeling document collections, presents a significant problem. Projection of such spaces onto a two-dimensional computer screen creates a confusing overlap of dimensions. This is alleviated to some extent by rotation and projection onto various subspaces. 3D rendering techniques may give a third dimension (a depth effect to the display), but require sophisticated software and hardware. It is also

a problem that the method is pixel-intensive—many pixels must be used for each object in order to achieve the depth-effect. Often we may encounter the problem of getting 'lost-in-space'—from where am I looking, in what direction? This is especially a problem when visualizing abstract, or inherently non-visual data. In contrast to architectural or natural objects, a visualization of abstract data provides the user with few clues to aid orientation.

Current information retrieval systems have many flaws that reduce their effectiveness. Chief among these is the lack of assistance the system provides to the user whose initial query failed to produce adequate response, or who wishes to expand the document search on the basis of the initial set of documents retrieved. Hypertext systems, through their explicit links, provide partial assistance in this. The visualization system presented here expands this assistance by including implicit links and providing a structured display of the relevant documents. By giving the user control over most aspects of the display, we provide a system that can be tailored to individual needs and perceptions. We feel that a document visualization system should be simple to use, especially as we must believe that such systems would be used as interfaces to traditional retrieval systems.

Desirable features for such a visualization system would be:

- *All documents should be presented, with their graphical representations, in one display.*

  The purpose of mapping inherently non-visual data from a world of text or numbers to a visual world is to give readers an overview of is to give readers an overview of the data, showing the organization and structure of the data set. It is the task of the visualization system, not the reader, to create this overview. Thus the user should avoid the problem of creating a holistic view from different displays. Note that presenting several windows with different perspectives simultaneously is not a violation of this principle. Such "small-multitudes" are an excellent method to present data (Tufte 1983, 1991).

- *The position and other graphical features of a document's representation, e.g., as an icon, should intuitively give information on a document.*

  Basically we use graphics to give an easy-to-read 'picture' of the data. While advanced display strategies may be useful for experts (e.g., for constructing wiring diagrams), we feel that most users will benefit from more intuitive displays, where the use of symbolic coding is kept to a minimum. However, we note that this principle will limit the possibilities for the positioning algorithms and transformations (e.g., rotation, projection, ...) that may be applied to a diagram.

- *The most important document attributes, as defined by the user, should be retained in the display.*

  There are several examples of how important aspects of the data may not be apparent from the use of statistical methods. An excellent demonstration

is given by F. J. Anscombe presents four very *different* data sets that have identical statistical measures, such as mean, standard error, regression line, t-value, etc. The problem is, of course, that statistical methods imply data reduction, which in this case results in an inability to distinguish among four different data sets. (Another problem is that statistical methods are impersonal, and do not necessarily reflect those aspects of the data that the individual user deems important.) In contrast, visualization techniques allow for presentation of large amounts of information—due to the high bandwidth of visual communication.

- *Users should be able to identify single documents, e.g., for retrieval of additional information.*

  Geographical information systems have taught us that a display may be used for *input* as well as *output*. By pointing at an object on the display additional information on this object may be presented. For example, detailed information—both textually or graphical—may be given on roads, buildings, etc., as we point at these objects in a map on the display. These systems introduce the idea of a *visual interface* to a database system. We may use such a feature to provide detailed information on documents, or document structure.

- *The display should be presented from the* user's perspective *and the user should be able to change the display interactively.*

  Scientific visualization utilizes the idea of interactive graphics, i.e., graphics that are controlled by the user. In order to understanding large or complicated data sets the user will have to go through a ding large or complicated data sets the user will have to go through a *process*, trying different visualization techniques, changing perspective, selecting new data attributes to be displayed, etc. Especially for viewing non-formalized data, such as document collections, it may be important that the user have the opportunity to view data from different perspectives. It is also important that the effects of any user-controlled change in the display be made visible, perhaps in a 'before and after' display mode. This will allow the user to apply a *what-if* strategy—introducing new parameters to study their effects on the display.

What we need is a visualization space where the user will be able to create and define the dimensions and determine the mapping of documents onto this space. The VIBE system is an effort in this direction.

3 The VIBE Approach

The VIBE (VIsualization By Example) system currently exists as two prototypes, one running on a Sun workstation under UNIX and X-Windows, and the other on a PC under Microsoft Windows 3.1. These prototypes work on a collection of

documents or data objects represented as flat files. Usually a document collection will be retrieved from current bibliographic database systems by giving a filter, the query. The documents are then displayed by VIBE in a virtual document space, established by letting the user create a coordinate system defined by *points-of-interests (POIs)* on the display. A POI defines a concept of interest to the user. If a text document collection is involved the POI may consist of a set of terms. (If the document collection is a set of database records with numerical or other fields, the POI may consist of field items.) The documents in the database are evaluated (scored) with respect to each of the POIs. The scoring mechanism is specifiable by the user and may be as simple as a frequency count of POI terms or a function on document attributes where quantitative data is involved. POI-terms may be weighted, and restricted to selected parts of a document (the term search will then only be performed in these parts). A POI may also be given a weight, to strengthen or reduce its influence.

Each point-of-interest (POI) is visually represented by a circle-icon on the display. Comparing documents to POI descriptions yields a vector for each document. All documents that meet minimum criteria (that get a score within a user defined range) are positioned on the display. The positioning of an icon is based on this vector of POI scores. Thus, the positions of a document show how the document *relates* to the POIs. As the system is used, the user will come to associate relative position with the contents of documents. The idea behind VIBE's positioning mechanism is that a document that matches a POI should be placed in the same position as the POI. A POI may thus be seen as a (simplified) example or prototype document that is given an example or prototype position.

In an information space with one POI only, all documents that are considered important will be positioned on top of this POI. VIBE will then perform as any non-visual document retrieval system (as the collection of icons on top of a POI may be presented sequentially, or as a list ordered by POI score). With two POIs, documents will be placed on the line between the POIs, in a position determined by the relative POI scores. However, the VIBE methodology will first become interesting with three or more POIs—allowing us to position objects with regard to more than one or two concepts/attributes.

A simple example of a VIBE display is given in Figure 1. Here we have three POIs, their icons presented as circles. In our example, the POIs are named: *document retrieval, hypertext* and *visualization.* Each of these is represented with a position on the display (circle-icon) and will be defined through one or more terms.

The positioning of four documents relative to these POIs are shown in Figure 1. Each document is represented by a rectangular icon. We find two document icons on the line between the *hypertext and visualization* POIs. Their positions tell us that the related documents are only influenced by these two POIs (if not, they would have been offset from the line). One document is more related to the first POI, another to the second. This is seen by the distance from the icon to the POIs. The icon positioned on top of the *hypertext* POI is, clearly, only influenced

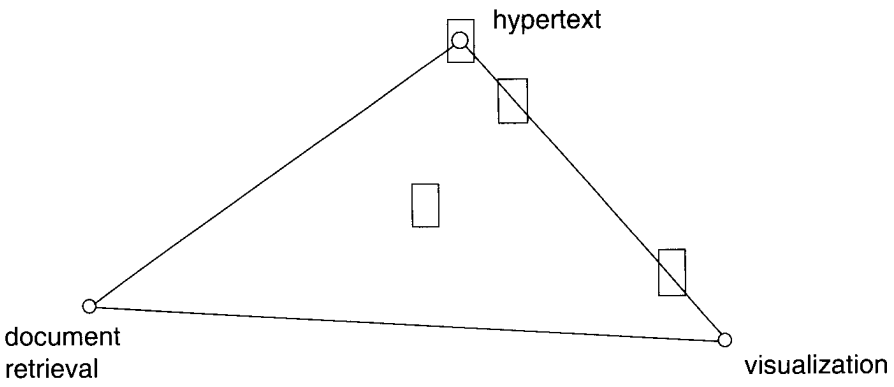

Figure 1: *A VIBE display*

by this POI. We also see an icon in the middle of the triangle defined by the three POIs. It's related document is influenced by all of the POIs. Since it seems to be positioned in the centroid of the triangle, the document is approximately equally related to each of these three POIs.

VIBE uses a simple algorithm for positioning documents, based on the ratios between POI scores. The positioning of an icon of a document D is based on the following input:

- The document score vector $D[d_1, d_2, \ldots, p_n]$, where n is the number of POIs, d_i the score of D on the respective POI and gives a quantitative value of the relevance of D to POI_i. Different methods may be used to determine d_i. For example, if POIs are defined through a list of terms, d_i may be set to the sum of frequencies for all keywords in POI_i on D. Since frequencies from POIs are to be compared, the d-values will be normalized with the average score for each POI (over all documents).

- The POI position vector $P[p_1, p_2, \ldots, p_n]$, where n is the number of POIs. Each p_i represents the display position (x, y) for POI_i.

These two vectors are combined into the sequence:

$$S = \{(d_1, p_1), (d_2, p_2, \ldots, (d_n, p_n)\}$$

1. If there are two or more elements in S with $d_i \neq 0$, two of these elements are removed from S. Each of these elements consist of a score value and a position—$(d_a, p_a), (d_b, p_b)$. A new element (d_i, p_i) is then created, based on these two elements. The score value d_i for this new element is $d_a + d_b$. The position p_i will be on the line between p_a and p_b, closer to the position that has the higher score. The distance from p_a to this intermediate position will be determined by $l_i = \frac{L \times d_b}{d_i}$, where L is the distance from p_a to p_b. The element (d_i, p_i) is then added to S, replacing the removed pair.

2. If there is only one element in S with a non-zero d-value, (d_a, p_a), the icon for D is positioned at p_a, if not, repeat from 1.

It is easily shown by mathematical induction that the final document position is independent of the order in which the elements from the set S is selected.

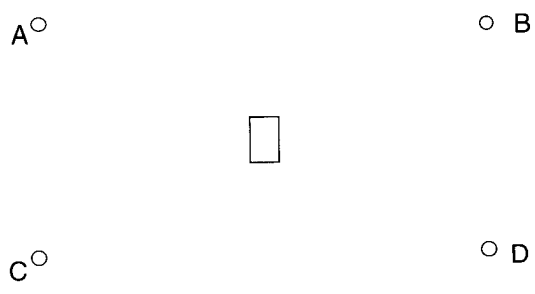

Figure 2: *A VIBE display with four POIs and one document*

Logically, every key word described in a POI will define an axis in the document space. Thus, VIBE transforms a virtual multidimensional coordinate system into a two-dimensional display. In practice, this implies that documents may be placed in the same position for several reasons. The coincidence of two documents may be real—they are identical with regard to the POIs, or false—resulting from the projected superposition of distinct locations. As an example, study the display in Figure 2. The document presented may be influenced, with equal score, by A&D, B&C or by A&B&C&D. However, by carefully positioning POIs, by moving POIs, etc., this problem may be controlled (see next section).

Icons can characterize or describe the documents they represent in many different ways. The attributes of the icons include its size, its color, and its shape. Of these attributes, only size is automatically applied in all cases. The size of an icon is an indication of the importance of the document it represents and is derived from the scores that a document has in relation to the specified POIs. For example, a document that has a low score on all POIs will be displayed as a small icon. Likewise, a document that has a high score on one or more POIs will be displayed as a larger icon. Thus, the documents shown as larger icons should be more closely related to one or more POIs than those having smaller icons.

4 VIBE User Interface

The VIBE system consists of two programs. A data definition program is used to calculate scores, based on textual and/or numerical data. Through a visual interface the user may define the record structure of any fixed or free format ASCII-file, and assign variable names to record attributes. Standard functions for

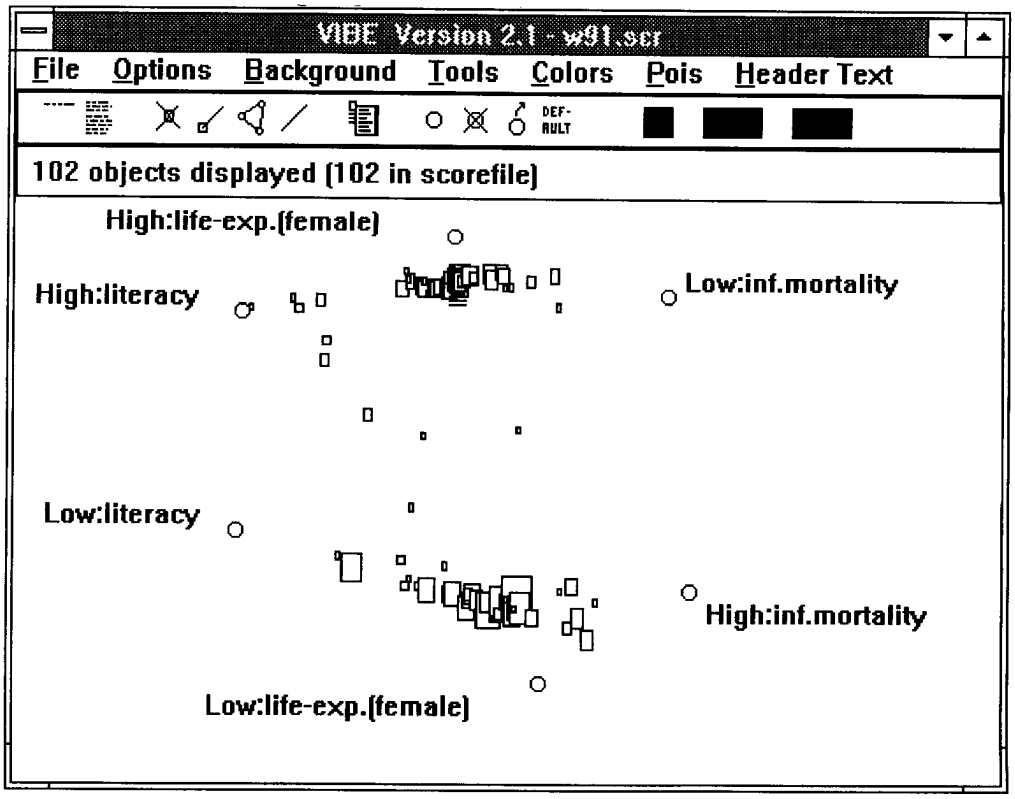

Figure 3: *VIBE user interface*

computing key-word frequencies, for normalization, etc., are offered, in addition to a formula language for doing score transformations.

The interface for the visualization program is shown in figure 3, with an example display. The example is based on a world-database, where countries of the world are positioned in an information space defined by *life expectancy, literacy* and*infant mortality.* The corresponding object attributes have been normalized, such that a higher than average score is attracted towards the 'high' POI, and a lower than average score towards the 'low' POI. The display tells us that most countries are in one of two groups, i.e., they either get a score on all 'good' POIs or on all 'bad' POIs.

The user may introduce new POIs to the display from the POI-menu. Through direct manipulation techniques POIs may be moved to a new position or removed from the display, or the system may be instructed to ignore the influence from certain POIs. Re displays after such changes are performed immediately, based on a table of POI component data kept in memory. Displays can be saved, retrieved individually, overlaid with other displays for comparison, 'pasted' into other applications or copied to a laser printer.

The influence from POIs on documents (the score values) may be visualized

by lines drawn from icons to POIs, the type of the line indicating the score value. Colors may also be applied to a POI. The color chosen will affect all documents influenced by this POI. Alternatively, a color may be assigned to special attribute values. For example, a color could have been used to present the common market countries in the display above.

By 'clicking' on an object VIBE will present all the information on the object that is available in the database. If more than one icon is positioned in the same position VIBE will present the topmost document—moving through the object stack for each 'button click'. Such an overlay of icons will be shown by lines under the icon, each line representing another object. A 'lens'-function will present information on all objects that are displayed in the current cursor position.

5 Explicit Links

As can be seen, VIBE will present the implicit links, i.e., the relationships between document contents, through positioning of document icons. Explicit links may be visualized through lines between documents or by color. However, the clustering of icons in VIBE displays makes these techniques difficult to apply. Colors are useful to represent group membership, e.g., all documents by a certain author, source, country, etc. However, the utilization of colors is limited by the fact that VIBE icons may overlap—making it difficult to recognize many different colors (Through experiments, we have found that a maximum of 8 colors may be used—the number being somewhat dependent on the degree of clustering in the display.)

As a general technique for visualizing links between documents, lines or arrows drawn between the respective document icons are probably the best option. Additional information on these links may be visualized through color, type and thickness of the line. For example, color may be used to distinguish types of links, each color representing a link category (hypertext links, references, citations, links between documents by the same author, ...). Line thickness and line type may be used to represent the strength of a link, perhaps defined as the number of links between two documents. A thin dotted line may be used to represent a weak link, whereas a thicker solid line may represent a stronger link. Further information on a link may be retrieved by 'clicking' on the line icon.

The VIBE system allows for group and split operations, where a collection of icons may be grouped into a 'super-icon'. The super-icon will be positioned in the centroid of the individual icons. To some extent this feature may be used to simplify diagrams, e.g., different sections of a document may be visualized as individual objects or as a collected object.

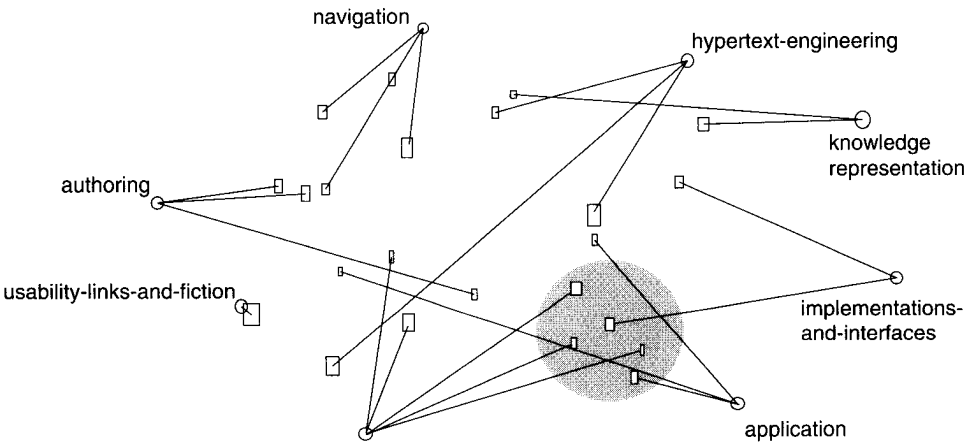

Figure 4: *VIBE display, links from section headings to section papers*

6 VIBE examples

An example of a VIBE display of a document collection is given in Figure 4. This collection consists of documents from the Hypertext'89 conference. Each document is represented by title and section headings (abstracts were not available for all papers). 26 of the 28 conference papers are displayed, i.e., 26 papers received a score above a minimum value in this eight POI information space.

Each POI name is a section heading from the conference proceedings, defined by one or more keywords. For example, the POI 'navigation' is defined with (stems of) the following keywords: *navigate, browse, path, traversal, tours* and*presentations.*

The structural information obtained from the organization of the proceedings, the explicit links, is visualized by lines from each heading (POI) to the papers presented under this heading. As seen, the structural information gives a somewhat different perspective than by the implicit links. For example, the five documents close together in the lower right corner of Figure 4 (indicated by a shaded circle) are dispersed over the conference sessions. In fact, the diagram tells us that the structuring of a conference and proceedings is not performed by studying contents of documents only, but that several different considerations have to be taken into account (three papers to a section, variation of speakers and institutions, etc.). This problem is also evident from the generality of some of the section headings (e.g., "usability, links and fiction", "applications"). We should therefore expect that documents get scores on more than one POI.

Since this is a collection of articles all from the same conference, there will not be references among the papers. In order to give an example of reference links, we have included a description of Jeff Conklin's "A Survey of Hypertext" (1987).

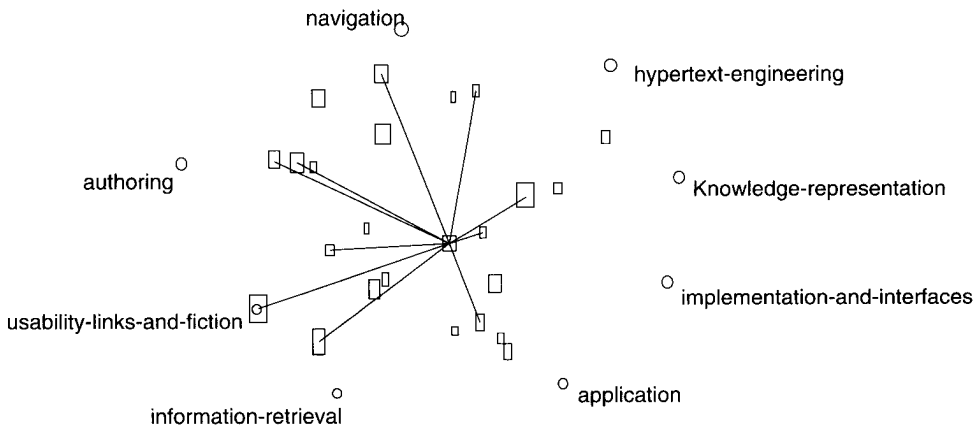

Figure 5: *VIBE display, references to Conklin's "A Survey of Hypertext"*

This general paper gets a score on all POIs, and is reasonably positioned in the middle of the diagram presented in Figure 5, where the links show the conference papers referencing Conklin's paper. From its center position we should expect this to be a paper that gives an overview of hypertext. This view seems to be supported by all the references to this paper, from documents in different subareas. Note that the lines here are document-document links, rather than document-POI links. If different types of links are used in the same diagram, e.g., session and reference links, these may be distinguished by the use of separate colors.

The same data is visualized in Figure 6, but now only with three POI's: *authoring, information retrieval* and *navigating.* The figure also shows an example of the presentation of additional information on a link, given after the user has 'clicked' on the link-icon (the line). This action has resulted in identifying both the referencing document and the referenced document, and has distinguished between these with an arrow pointing to the referenced document. A total of 24 of the conference documents get a score above a minimum value on these POIs. Two of these fall between authoring and navigation, four between navigation and information retrieval and two between information retrieval and authoring. The rest of the conference documents fall on top of the POIs. Only one paper, in addition to Conklin's, gets a score on all the three POIs. It seems that these POIs are good discriminators for at least 24 of the 28 conference papers.

One should note that VIBE is a *visual interface* to a document collection. Using VIBE is therefore a dynamic process, and hard copies of VIBE displays, as presented in Figures 4 to 6, are only snapshots of this process.

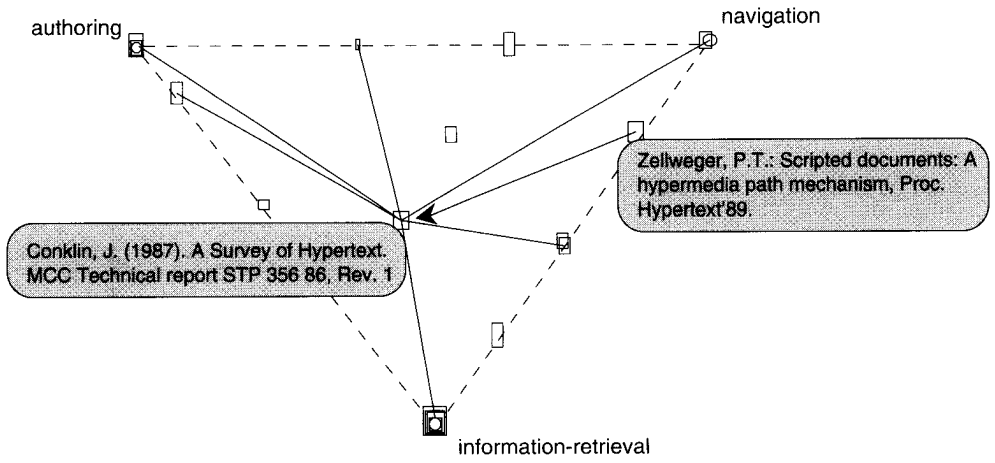

Figure 6: *VIBE display with three active POIs, references to Conklin's "A Survey of Hypertext"*

7 Applications

The UNIX-prototype version of VIBE has been used extensively for a research project funded by the Office of Scientific and Technical Information (OSTI), Department of Energy (DOE). This research project investigated methods of extracting structural information from large scientific bibliographic databases, information that can not be extracted using the traditional Boolean search mechanisms of such systems (Sochats *et al.* 1991). In this project, VIBE was used on a sample collection of documents from this database on a very specialized and constrained subject area—*inertial confinement.* Each document reference in this database consists of a full abstract and several indexing terms organized in a hierarchy, together with other information on documents (title, author names, country of origin, etc.). The indexing has been performed manually, by professional indexers.

Figure 7 shows an example display from this application, where documents are displayed in a 3-POI information space, defined through the hierarchy *laser, target*and*ion-beam.* As expected, most documents fall on the line between the POIs, indicating that indexers has managed to apply this hierarchical term structure correctly. The three documents that fall between the POIs are all general articles on the field. Note that overlaying icons (same position and size) are visualized by a line under the icon of the first document.

On the basis of these data, an explicit link structure has been overlaid on diagrams of the type shown in Figure 7. The overlay structures have included links between papers with the same authors, papers from the same institutions

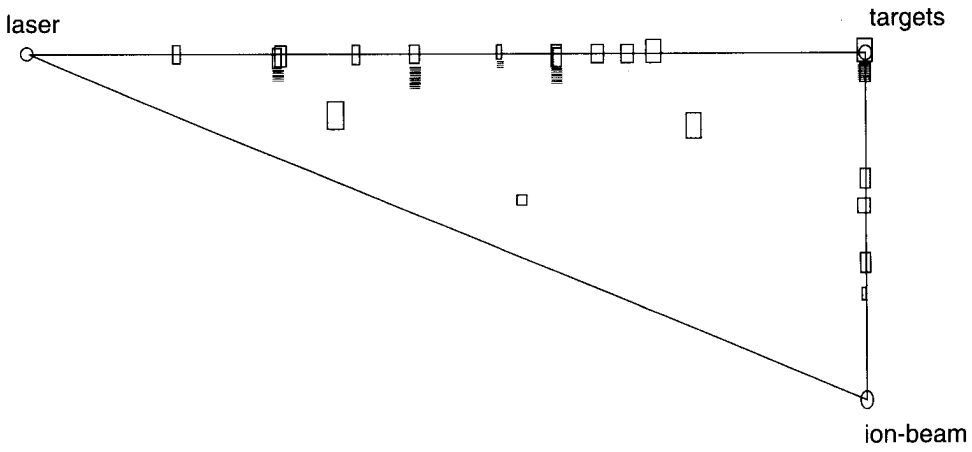

Figure 7: *VIBE application (DOE Energy database)*

and papers from the same countries. Different graphical features, from links shown as lines or the use of colored icons have been applied.

Another test application for VIBE is to provide an alternative view of a personal archive, e.g., positioning document files in a user-oriented information space instead of the common directory structure. The file structure could be visualized through group identifiers (color) or through lines between documents. We have also used VIBE as an e-mail organizing tool, where messages are positioned according to their content and where links show the chronological order of the messages (original message to replies, etc.). The problem with the latter application has been that e-mail messages are often extremely context dependent—making it difficult to position the message from its contents alone (see next section).

8 Discussion

The uncertainty associated with natural language is a problem for all systems that work directly on the document text. Users approach an information retrieval system in the hope that the documents retrieved are meaningfully related to their information needs. However, current information retrieval systems cannot match documents and information needs semantically. Rather, the retrieval is based on a lexical match between the query and the documents. Thus the effectiveness of the retrieval will depend on how precisely one can describe concepts in queries and documents.

The nature of the documents themselves, and the data chosen to represent the documents, also set limitations. For example, one gets a different picture of the Hypertext'89 conference if full text is used, compared to titles and chapter

headings. Further, the document collection sets practical restrictions to how closely one can focus in on the documents.

In most information retrieval systems the limitations of natural language are hidden from the user, even though they affect the output. In VIBE they become directly apparent (documents scattered over the information space, documents not related to POIs, etc.) and the user can take action to counter them.

Visualizing explicit links on top of these diagrams may give important additional information to the user, for example:

- A user (reader) may find a cluster of interesting documents in one location and may follow references to other areas to find additional documents of interest.
- Reference links between documents in a cluster may strengthen the relationships among these.
- Reference links from clustered documents in quite different directions may weaken the similarities between these documents (see for example Figure 4).

The possibility of combining implicit and explicit links, by turning implicit links into explicit, has been suggested. For example, the user may discover, through inspection of individual nodes or documents, that documents may be organized into different groups. This may be achieved by connecting icons, by drawing lines between the various icons or by using some 'lasso' mechanism. Additional information on such links may be added by the user, e.g., by describing what they have in common.

Retrieving information on explicit links is a problem. Few bibliographic databases provide such information. However, this problem may be overcome in the future as the extensive usage of word processing ensures that more documents will be available in electronic form. Modern storage techniques, high capacity networks and standards for representing and communicating documents will greatly enhance the opportunity of retrieving full text of documents or at least comprehensive abstracts from bibliographic databases. Explicit links, as citations, may then be extracted from these data by automatic or semiautomatic techniques.

VIBE visualized explicit links as lines between documents, while the documents themselves are presented as small rectangular icons. Care has been taken to use as few pixels as possible, in order to avoid a clustered screen. This is important, as the idea with using visualization is to give an overview over a data collection. However, even with our 'pixel-meager' methods it is difficult to avoid cluttered displays with large numbers of documents and links. For such large collections (from experience, collections of several hundred documents) we have been forced to introduce data reduction methods, such as filtering, zooming and

a system where groups of individual documents may be aggregated and represented with a group-icon. The idea is to be able to simplify the display, with the least possible data reduction.

9 Conclusion

We have presented a system for visualizing a collection of documents and the links between these, as represented by references, citations, hypertext links, etc. A central idea behind the system is that semantic similarities will be presented through the positioning of document icons in a user-defined information space. The explicit links between documents, will be visualized by VIBE on top of this view.

In theory, this visualization strategy may be used to present intra-document links as well as inter-document links. We may envision VIBE used to position hypertext nodes in a user defined space, and presenting the hypertext links as a flat structure on top of these nodes. For an author VIBE may provide a help to organize the nodes, while a reader may use the system to get different perspectives on the data. In this respect, we may think of VIBE as an option in a hypertext system, allowing the user to get an overview of the node collection—in a user-defined information space.

However, as VIBE is based on the processing of natural language, it will need quite extensive information in order to place nodes in their right positions (as index terms, abstracts, etc.). If nodes are textually small, the text unstructured or context dependent, we may find that the positioning of nodes, at least with regard to the POIs as concepts, will be more at random.

10 Future Work

The VIBE system is a prototype. It has been used for several test cases, but for few real applications. Results are promising, but the primary aims for further work are to perform a validation of the methodology. A 'production' version of the system (PC only) will be used extensively, both by us and by others, to test VIBE on different type of document collections. An important task will be to compare VIBE with alternative methods, as statistics, tables and more traditional visualization methods. Through this process we will try to determine the usability of VIBE diagrams and their application dependent limitations. An important question, that we will try to answer through these tests, is to determine to what extent VIBE will give the researches information that would not have been possible to obtain by use of more traditional methods.

References

Anscombe, F. J., (1973). Graphs in Statistical Analysis. *American Statistican*, 27:17–21.

Conklin, J., (1987). *A survey of Hypertext.* MCC Technical report STP-356-86, Rev. 1, Microelectronics and Computer Technology Corporation.

Doi, A., M. Aono, N. Urano & K. Sugimoto, (1991). Data visualization using a general-purpose renderer. *IBM Journal of Research and Development*, 35(1/2):45–57.

Fairchild, K. M., S. E. Poltrock, & G. W. Furnas, (1988). SemNet: Three-dimensional graphic representations of large knowledge bases. In: R. Guindon, editor. *Cognitive Science and its Applications for Human-Computer Interaction.*, Lawrence Erlbaum, Hillsdale, N. J.

Glushko, R. J., (1989). Design Issues for Multi-Document Hypertexts. *Hypertext'89 Proceedings.* ACM, New York. Pages 51–60.

Sochats, K. M., M. Weiss & J. G. Williams, (1991). *Intelligence in Large Scientific Databases.* Report to the Office of Scientific and Technical Information, Department of Energy, School of Library and Information Science, University of Pittsburgh..

Tufte, E. R., (1983). *The Visual Display of Quantitative Information.* Graphics Press, Connecticut.

Tufte, E. R., (1990). *Envisioning Information.* Graphics Press, Connecticut.

van Rijsbergen, C. J., (1979). *Information Retrieval,* Second Edition. Butterworths, London.

Young, F. W. & P. Rheingans, (1991). Visualizing structure in high-dimensional multivariate data. *IBM Journal of Research and Development*, 35 (1/2), 97–107.