COVER PAGE

Empirical Study of a 3D Visualization for Information Retrieval Tasks

Gregory B. Newby

University of North Carolina at Chapel Hill

Author contact information:

Dr. Gregory B. Newby
School of Information and Library Science
University of North Carolina at Chapel Hill
CB 3360 Manning Hall
Chapel Hill, NC, 27599-3360
Email: gbnewby@ils.unc.edu
Fax: 919-962-8071
Phone: 919-962-8064

Suggested keywords:

Information visualization
Information retrieval
Visual interfaces
Evaluation

Empirical Study of a 3D Visualization for Information Retrieval Tasks

Gregory B. Newby

University of North Carolina at Chapel Hill

There are many challenges to visualizing information including choosing between 2D and 3D interfaces, navigation and interaction methods, and selecting an appropriate level of detail. Visualizing information retrieval (IR) search results, including Web search engine results, poses additional challenges, notably the determination of appropriate relative locations for terms and document in a visual display. Latent Semantic Indexing and related techniques offer approaches to visualizing relations among terms and documents. In this work, "information space" is presented as a framework for discussing relations among terms and documents, and a technique related to LSI is utilized to generate information spaces from IR search results. This paper provides an overview of more than three decades of work on information visualization, identifying several trends and some relatively unexplored areas. A experimental evaluation of a prototype interface for visualizing IR results is described. Results indicate that the 3D navigation interface for IR search results was usable, but that subjects had difficulty with some aspects. Further study and development of 2D and 3D methods for interacting with retrieval search results is suggested.

Introduction

Visualization for teaching and learning is nearly ubiquitous. The use of still and moving images in education ranges from diagrams, drawings and photos in textbooks, to the construction of models, to computer-based systems for navigating in virtual worlds. In many cases, these visualizations represent either reality, or an approximation of a physical reality. For example, photos or models may depict a scene as it may have happened in history, or a movie can tell a story of events.

In other cases, visualization might be used to help understand data that do not have an obvious tangible physical manifestation. For example, we might use pie charts to describe the allocation of a budget, or data flow diagrams to describe the processing in a computer program. It is this second group of cases, without a compelling or obvious scheme for going from a set of data to a visual representation of those data, that this work is concerned. In particular, we will address the visualization of relations among relatively unstructured textual documents and the terms within them.

This paper describes a method for visually displaying and interactively navigating among results from an information retrieval (IR) system. IR systems may contain thousands, millions or (in the case of the Web) billions of textual documents. People pose queries to these IR systems (for the Web, the systems are typically called search engines). The response sets to queries may consist of dozens to many thousands of documents.

We propose visualization of relations among documents and the terms and concepts they contain, in order to effectively identify useful documents from a retrieved set of candidates. Instead of wading through a long list of "hits" (that is, potentially useful documents) ordered by a retrieval equation, visualization can serve to cluster similar documents and identify regions of potential interest. Visualization can also be used for iterating the search process, narrowing in on areas of potential value.

We view visualization of IR results as a short and intense education experience. In order to find useful information, an information seeker must quickly assess the value of the hits returned by an IR system and decide whether to view any documents, re-phrase the search, or abandon it. If truly useful visualization schemes for text are introduced, we hope to make visualization of IR results as useful as, for example, visualization of fluid dynamics data. Most people have seen visualizations of air flow over an airplane's wing, or of gas expansion during ignition in an engine, or other examples of simulations of dynamic systems – without visualization, how could we find value in the millions or billions of individual data points?

Visualization can be powerful for many purposes. Tufte (1990), for example, has demonstrated the utility of visualization for abstract mathematical concepts such as spaces of 4 or more dimensions. Shneiderman (1997) has presented extensive treatment of the power of visualization and methods for designing and evaluating effective data visualization systems. More recently, Card et al. (1999) have written a thorough treatment of the importance and difficulty of information visualization. For information retrieval (IR), there has been a long-standing interest in visualization of documents, collections and retrieval results (see Newby, 1993 for an historical review).

Below, we will discuss the role of visualization, especially navigable interactive visualization, for IR. Rather than investigating the role of visualization as the primary interface to an IR system, we will consider visualization of search results. An analysis of related literature indicates that the challenges of visualizing the relatively unstructured text found in IR system results are somewhat different from other domains for visualization. These challenges have not yet been extensively addressed.

Overview of Yavi and Information Space

For IR and for human-computer interaction in general, the role of visualization, navigation and 2D or 3D interfaces is subject to debate. Scholars have discussed whether 2D alone is sufficient, and whether visual interfaces for IR can be more effective than text-only interfaces. This research will attempt to contribute to the discussion by reviewing related research then providing analysis of empirical data from analysis of a prototype system.

To determine whether a simple 3D navigable visualization system for retrieval results was perceived as usable by trained searchers, an evaluation the "Yavi" prototype was made. Yavi is a prototype 3D application for exploration of an information space where similarity relations among terms and documents may be viewed, and documents may be retrieved (see Figure 1). The outcome suggests that 3D views may be more confusing than 2D, but that subjects were able to use Yavi to make judgments about document and term relations in a visualization of simulated IR search results.

Information Space is defined here as the *set of relations among items held by an information system* (cf. Ingwersen, 1996). We may think of a collection of documents and their related terms (words) as composing an information space, so long as relations among the documents and terms are measured. Perhaps the best known model of information space for IR is the vector space and its progeny (Salton & McGill, 1983). Notice that information space does not need to have a clear visual interpretation – indeed, the lack of an obvious visual interpretation of most IR information spaces is an important topic addressed below.

Building an information space for Yavi is closely related to Latent Semantic Indexing (LSI). For Yavi, an information space is based on a numeric similarity score for each pair of terms found in a set of retrieval results. A term-by-term co-occurrence matrix for the retrieval results is submitted to an eigensystems analysis to reduce the dimensionality of the matrix. Term vectors in the resulting eigenvector matrix may be thought of as geometric coordinates in a multidimensional space. This is similar to the standard vector space model, except that the term vectors are not mutually unrelated (orthogonal). Documents from the collection are placed at the center of the terms they contain. This process is not extensively described here; the reader is instead referred to Deerwester et al. (1991), Kruskal & Wish (1978) and Newby (1997). Yavi is, essentially, a visualization tool for multidimensional Information Spaces consisting of terms and documents found in retrieval results.

The next section provides a more descriptive basis for our investigations of information space, and presents the long-term goals of the current research program. The following section reviews literature related to visualization tools for retrieval systems, and explains some of the difficulty in applying visual techniques to unstructured text. The remainder of the paper describes a usability study of the Yavi prototype. Fifteen respondents completed two tasks each with the system, and responded to a pen-and-paper questionnaire about the tasks.

[Insert Figure 1 here. Note to editors: Figure 1 is 4 separate images. They are included in this Word document, and as standalone files. Color images are included for possible electronic publication; the publisher should use black-and-white versions for print reproduction. Image files accompany this document, and are named "gbn-fig1a-bw.tif" and so forth.]

Contemporary Information Visualization

The use of a visual metaphor for interacting with computers has helped to make visualization ubiquitous, and to provide computer users with a facility for directly manipulating intangible objects. However, the problems identified by early non-computer-based information visualization systems remain. Specifically, how do we decide how to relate items graphically, and how should people interact with these information items?

A reasonable approach to addressing these problems is to provide a depiction of a well-known physical domain as the basis for the organization of information objects. Pejtersen (1992) provides one of the best-known examples with her "Book House." This system is intended as an extension of the library card catalog, with facilities for browsing virtual library shelves, getting book recommendations, and perusing different "rooms" in a virtual library. The problem of manually classifying or locating items (books) is alleviated by the pre-existing cataloging scheme. Extending this model to items that have not been cataloged, such as a collection of Web pages, could be problematic (and also stretch the limits of the physical library-as-metaphor).

Stuart Card, Peter Pirolli, Jock Mackinlay and their colleagues (Card et al., 1991; Robertson et al., 1991; Pirolli, 1995) have addressed the role of visualization of large and relatively complex data collections. One example is their hyperbolic tree. This is a visual interface to the organizational hierarchy in an organization. The hierarchy is represented in a tree structure, where clicking on a "branch" brings up information about an individual or sub-organization. Another example is their information wall, in which a large series of relations among items (such as in a programming flow chart) may be easily visualized by providing detail for a central region, but less detail in other regions.

Tamara Munzner (1998) has developed a series of systems for visualizing graphs and other link-structured data, including the World Wide Web. She uses a hyperbolic space, enabling a more efficient representation of complex data. Her work includes visualization of lexical thesaurus data (Munzner et al., 1999).

Work by Johnson (Cochrane & Johnson, 1996) sought to visualize thesaurus information. Broader, narrower and related terms were represented visually in a tree similar to Munzner's hyperbolic maps. Documents from a collection could also be mapped visually, based on the thesaurus terms from their bibliographic data.

Xia Lin has worked to visualize relations among data items, including individual scientists, Web-database content and subject thesauri (e.g., Lin, 1999). He has used a mixture of techniques, including Kohonen self-organizing maps (SOM, see Kohonen et al., 1999), term co-occurrence patterns, and categorical relations in thesauri (such as broader/narrower terms).

A project to generate a map of multiple academic domains was carried out by Chen and his colleagues (Chen et al., 1998). The resulting map, similar to a Kohonen self-organizing map, identified relations among domains based on term occurrence patterns in documents that belong to the domains. Like the subject maps of Kohonen, Lin and others, Chen et al.'s map could be used to visually present major topical groupings of a domain in a 2 dimensional grid.

Several projects have resulted in interfaces or graphical depictions of a document collection with similar qualities to Yavi. Earlier work includes Lyberworld (Hemmje et al., 1994), Bead (Chalmers & Chitson, 1992) and Vibe (Korfhage, 1997). These three systems presented a visualization of term and document relations for IR, based on patterns of term occurrence within a collection. The Information Visualization Group at the Pacific National Laboratory has produced several prototypes as part of their SPIRE system (see http://www.pnl.gov/infoviz/technologies.html) geared towards large-scale data mining, as well as IR. Their "Galaxies" and "Themeview" interfaces both present a navigable visual depiction of relations among concepts from a document collection.

Most of the information visualization techniques presented in this section are either non-interactive (that is, simply a visual presentation of a set of data), or interactive but not a gateway to further information or transformations. A desirable feature for visual information retrieval interfaces is for the visualization to be a means to actually retrieve interesting documents.

Visual interaction with large data sets may benefit from Shneiderman's "visualization mantra" (Shneiderman, 1996):

1. Overview first
2. Zoom and filter
3. Details on demand

In his work, Shneiderman has emphasized the role of visualization for data that would otherwise be difficult to understand, such as for seeking an apartment in a metropolitan area based on multiple criteria, or selecting a movie based on awards received or nominated. A characteristic of Shneiderman's work is that multiple views of the same dataset may be taken in order to facilitate understanding.

Unlike unstructured information retrieval, the data Shneiderman and his colleagues most frequently work with have multiple pre-defined data types. (For example, apartment seekers might be interested in geographical location, crime rate, cost, number of bedrooms, and whether pets are permitted.) Most IR systems, however, are able to act only on the presence or degree of a similarity relationship. Aspects of IR collections that may fit Shneiderman's style of visualization might include author overlap, catalog terms, length or type of item, and so forth. The main limitation of some of the techniques demonstrated by Shneiderman for IR has to do with real-world performance: whereas most of Shneiderman's tools operate on hundreds to tens of thousands of data points, text retrieval systems often deal with hundreds of thousands or millions of items.

Evaluation of IR Visualization Systems

One of the challenges in evaluating the effectiveness of methods for visualizing information retrieval results or other approaches to IR visualization is the ubiquity of existing visual metaphors. Consider a typical Web search engine: people use a visual metaphor for their core system interaction (that is, manipulating a mouse to select fields for data entry and submit a query for processing). However, search results are typically presented as a (non-graphical) list of possibly relevant documents. While the visual metaphor is in effect for the underlying human-computer interaction, the actual retrieval results are presented as linear text, supported by some hyperlinks.

A recent example of visual IR system evaluation may be found in Sebrechts et al. (1999). They were interested in comparing three types of information retrieval interfaces: text, 2D and 3D. Their evaluation also compared novice to professional searchers, and examined the impact of experience. A controlled experimental setting was augmented by the opportunity for subjects to provide qualitative feedback about their search experiences.

Sebrechts et al. found there to be a high "interface cost" for the 3D system, in that it was more difficult to learn and to use than the other systems. They determined that different types of search tasks were better suited to text, 2D or 3D interfaces. To generate their findings, they needed functional systems for text, 2D and 3D IR, a series of search tasks, and a number of research subjects to perform the tasks. Statistical significance of results was aided by the relatively large number of search tasks: for each of 6 search topics (such as “smoking bans” and “gene therapy”), subjects were asked to complete 16 tasks such as locating, comparing and describing documents or clusters.

Sebrechts et al. found there to be an interaction between the type of task and the type of interface, as well as a role for searcher experience. This finding is consistent with that of Allen (2000), mentioned above. There was no single recommendation for the use of 3D over 2D over text, but instead a series of situations in which each type of system might be best.

Another example of visual IR system evaluation is Swan and Allan (1998). They compared a text system, a GUI system and a 3D system. They were interested in the ability of subjects to find documents covering various aspects of a search topic. Although the results indicated a slight performance advantage for the 3D interface, the overall effect was ambiguous. Whereas some subjects were enthusiastic about the 3D interface, some were not – particularly more experienced searchers.

According to Sebrechts et al., there have been, “very few studies directly comparing the effectiveness of visualization against functionally equivalent traditional interfaces for IR applications. Previous data shows [sic] only modest benefits of visualization in this context (1999, p. 4).”

In addition to the comparison of functionally equivalent 3D and non-3D interfaces, we can identify at least one other type of IR visualization evaluation: one which attempts to assess the role of a visualization system without comparing it to a more traditional system. For example, Nowell et al. (1996) evaluated a 2D display of document attributes without comparison to a text system.

Dillon (2000) offers a model of incorporating individual differences into information system design by attending to the semantic relations among information items. Although his terminology is different from what is used here, his central argument is similar: it is necessary to understand how individuals perceive items in an information space in order to decide how best to present them. He offers the perspective that evaluation of information visualization systems must take into account individual differences, in addition to focusing on differences among system or task types.

Overall, evaluation of visual interfaces for IR can be challenging:

- Generating prototype systems suitable for evaluation is time-consuming, and important features may be overlooked.
- While IR systems typically store many thousands or millions of documents, this scale is seldom achieved in prototype visualization systems.
- Human subjects are likely to be familiar with non-visual IR interfaces, but not visual interfaces. This could put the visual interface at a disadvantage, or create a need for extensive training.
- The benefits of human information navigation might not match usual IR performance measures such as recall and precision. For example, an information visualization might provide a better overview of a subject domain than a text based system, but this aspect might not be measured during an evaluation. In

particular, visualization may be valuable for educating about an IR domain, yet this role does not fit well with typical relevance-based IR evaluation.

What do we know about IR Visualization?

From our analysis of information visualization, we conclude the following:

- The use of visualization techniques and metaphors underlies modern human-computer interaction;
- Information visualization for various domains has demonstrated utility and many advocates;
- Most tools for information visualization have these limitations:
    - Not suitable for large collections (but may be suitable for visualizing subsets of large collections);
    - Automatic means for assessing relations are based on statistical properties of term and/or document occurrence data;
    - Sophisticated methods for structured data (such as hierarchies or thesauri) do not have clear applicability to relatively unstructured text such as Web pages or journal articles when automated processing is required.

Visualizing Information Space

Information retrieval typically focuses on the matching of information needs to documents (or document surrogates such as abstracts or citations). Matching is based primarily on topicality, which is considered an important component of relevance (Schamber, 1994). The main underlying challenge in IR is to base this match only on terms found in the document and the query (statement of information need), with relatively little help from other sources.

Other potential sources beyond the topic might include characteristics of the information seeker, such as standing profiles of information need (Hull, 1999), knowledge of the information seeker's situation (Schamber et al., 1991), and individual differences among seekers (Borgman, 1989). Further sources of help in matching the query to desirable documents can include controlled vocabularies or thesauri, or markup or other structure in texts (Hawking et al., 1999).

In spite of these possible sources of help, the dominant basis for retrieval systems remains the terms in queries and documents, along with knowledge of the terms in other documents from the collection. Three general approaches to this type of matching are prominent in the IR research literature: Boolean retrieval, probabilistic retrieval and vector retrieval.

Boolean retrieval is the process of matching documents to query terms based on whether terms are present in each. This is the technique preferred by online library catalogs, database vendors such as Dialog and Westlaw, and others. When a document contains none of the query terms, the relation to the query is the empty set (that is, zero or no relationship).

In contrast, probabilistic retrieval and vector IR (see Salton & McGill, 1983) make use of patterns of term frequency in documents in order to outperform Boolean retrieval. For both probabilistic and vector techniques, equations relating terms and queries take into account the relative weight of terms within the documents and queries. In order to identify documents with similar topics but fewer matching terms, query expansion may be applied to add interesting terms to the original query. Even so, the fundamental process of probabilistic and vector IR is to first determine which documents in a collection contain the (expanded) query terms, and then to rank the appropriateness of each candidate document to the query based on an equation. The choice of weighting formulas and similarity equations is of paramount importance to these types of systems.

Another approach to information retrieval seeks to enhance the likelihood that documents with similar topics to a query may be retrieved when there are few or no matching terms. Latent Semantic Indexing (LSI, see Deerwester et al., 1990) starts with the same terms and documents that Boolean, vector and probabilistic methods do. Then the term by document matrix (i.e., the inverted index) is subjected to a mathematical transformation, the singular value decomposition (SVD, see Golub & van Loan, 1996).

The SVD removes redundant information in the matrix and organizes the remaining information. The outcome is a series of new matrices that identify the relations among terms and the relations among documents. Essentially, LSI considers a term as consisting of the sum of some portion (weight) of all the other terms in a collection. A document consists of weights from other documents, or alternatively the weights from its terms.

According to the LSI model, Boolean-style retrieval is not appropriate. Instead, the basic approach to matching a query with documents is to determine the best-matching documents regardless of which terms they contain. For example, the SVD may determine that four terms have very similar patterns of occurrence with other terms across the collection: pigs, swine, boar and hogs. Thus, a query with one term (such as “pigs”) could have a high similarity with a document that has a related term (“hogs”). Rather than expanding queries based only a small set of term relations, LSI considers all terms potentially related to each other, and all documents to be similarly related.

A visualization system based on LSI is described in Chen (2000). He combined LSI processing on a relatively small corpus of 169 ACM SIG/CHI articles with Pathfinder analysis (Schvaneveldt et al., 1989). The outcome was reminiscent of Card’s hierarchical organizational charts mentioned earlier, but with clustering based on conceptual relatedness. Chen found few statistically significant results (his sample size was $n=6$), but did identify a strong positive correlation between recall scores and subjects’ associative memory test scores ($r=0.885$, $p < 0.001$).

The Yavi visualization system under investigation here is based on an IR approach called Information Space, which is similar to LSI. Unlike Chen’s (2000) hierarchical display of concept relations, Yavi uses a point-cloud display, where terms and documents are represented as points in a 3D space.

The main differences between how Yavi’s information space is built and generic LSI are pragmatic:

- Information Space starts with the term by term correlation matrix while LSI starts with the term by document occurrence matrix;
- Information Space performs eigensystems analysis while LSI performs a Singular Value Decomposition (SVD);
- Information Space has not made use of eigenvalues for scaling document vectors while LSI does make use of singular values (square roots of eigenvalues);
- The Information Space approach does not assume that higher-dimensioned eigenvectors are without merit while LSI historically has sought to eliminate higher-dimensioned eigenvectors

Further details may be found in Newby (1997). For visualizing information retrieval, the role of the Information Space technique is providing a measurement of the relationships among terms and documents. Because of the prioritizing effect of the SVD and eigensystems analysis, a large matrix may be visualized using only a few dimensions.

For example, the input to an Information Space eigensystem problem may be a matrix of thousands of terms occurring in hundreds of thousands of documents. The resulting eigensystem removes the redundant information from the original input matrix, and presents a new matrix where the first 3 or 4 columns can account for 15 to 25% of the information from the original larger matrix.

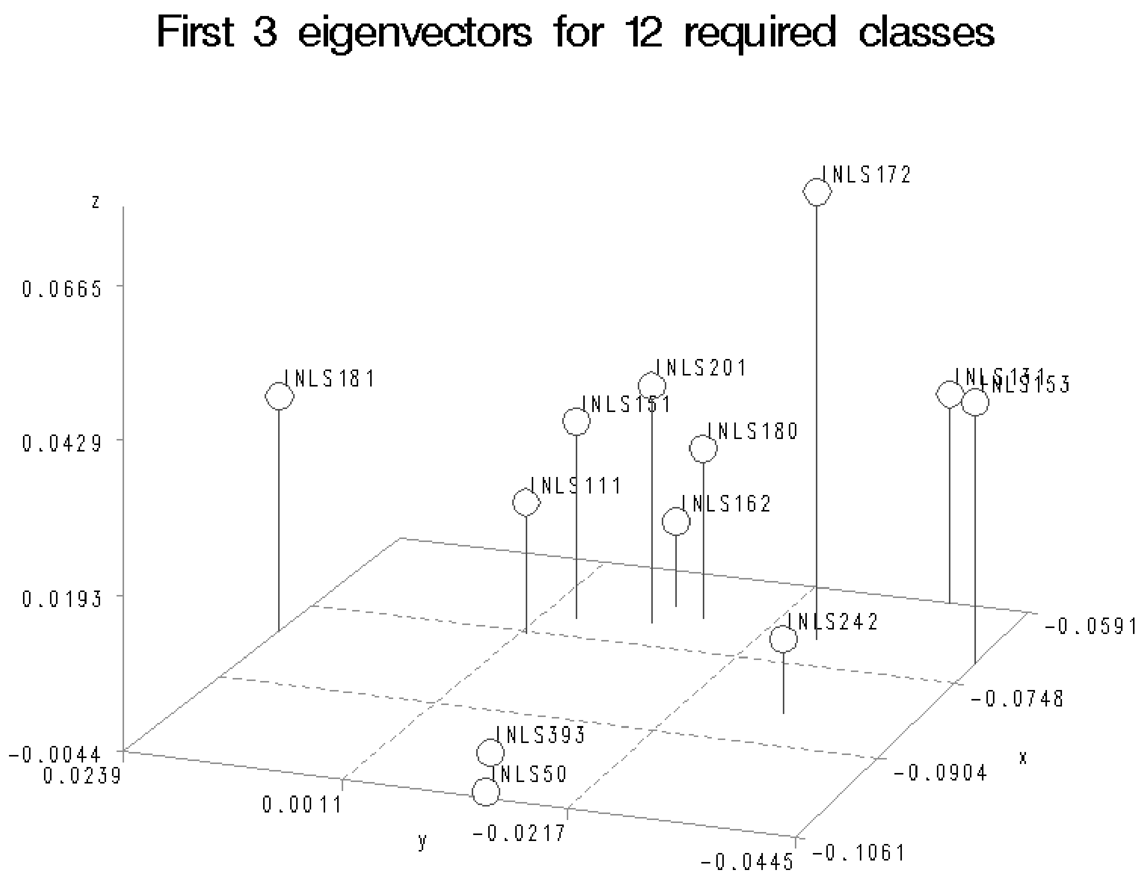
Figure 2 is a 3D scatter plot of the first three dimensions of such a process. The depiction is of relative locations for 12 courses, as discussed under the Methodology for COURSES below. It may be easier to distinguish relative document locations in Figure 2 than in Figure 1, however Figure 1 is essentially the same type of material, except in a 3D navigable display.

The specific process to produce the locations for these 12 courses was:

1. Identify interesting terms from the course catalog descriptions. This resulted in 111 terms.
2. Build a term-by-term co-occurrence matrix by counting the number of documents in which each term co-occurs with each other term.
3. Build a term-by-term Pearson product moment correlation matrix from the co-occurrence matrix. This measures the extent to which terms have similar patterns of occurrence to other terms.
4. Perform an eigensystems analysis on the correlation matrix.
5. Figure 2 displays the 1^{st}, 2^{nd} and 3^{rd} eigenvectors of the correlation matrix.

With these particular data, the combined variance accounted for by the first through third eigenvectors is 34%. In other words, 34% of the information in the original correlation matrix may be visually presented in 3D. This is the basis for the visualization. Although 34% is not as high as we might like, respondents agreed that there is some value in the 3D presentation. With Yavi, functionality is provided to view any 3 contiguous eigenvectors.

Figure 2: 12 Required Classes in Information Space



Key:

inls 50 Introduction to Computing

inls 111 Information Resources and Services

inls 131 Management of Information Agencies

inls 150 Organization of Information

inls 151 Organization of Materials I

inls 153 Resource Selection and Evaluation

inls 162 Systems Analysis

inls 172 Information Retrieval

inls 180 Human Information Interactions

inls 181 Internet Applications

inls 201 Research Methods

inls 242 Curriculum Issues and the School Librarian

inls 393 Master's Paper

Contrast this ability to visualize Information Space relations to the abilities of Boolean, vector or probabilistic systems: for all three, the assumption is that terms are unrelated (orthogonal). This means that a collection with 10,000 terms would require 10,000 term vectors (or dimensions) to represent. There is no obvious mapping of such a term space into two or three dimensions for visualization or navigation. Even the incorporation of term weights (which specify relations among terms) does not lead to a clear visual interpretation.

Thus, in the Information Space technique we see potential solutions to some of the challenges to information visualization identified earlier:

1. Information Space is useful for unstructured text as it occurs in documents, Web pages, etc.;
2. An approximate visualization of a document collection is possible; and
3. Information retrieval using Information Space and LSI have been found to be comparable in performance to vector or probabilistic systems (e.g., in the TREC conferences).

In the rest of this paper, an implementation of an Information Space visualization tool, Yavi, is described and evaluated. For this research, very small domains were investigated – smaller than would be found in an operational IR system. However, the Yavi system is usable as an interface to a full-scale IR system based on Information Space, in which queries are matched against hundreds of thousands of candidate documents. Then, a subset of candidate documents, along with all query terms and related terms, is presented visually for navigation-based retrieval using Yavi.

Methodology

Fifteen respondents participated in the evaluation. Each was paid $10/hour for their involvement, which included two additional research projects not reported here. Most (12) respondents attended an introductory training session conducted by the investigator where major Yavi features were discussed and cognitive tests were administered. The respondents were able to practice with Yavi for as long as they liked with a sample task to demonstrate Yavi functionality.

After training, respondents were given a packet of instructions and survey forms for completion at their leisure. Respondents needed to use two separate Information Spaces to complete tasks and answer questions about the information spaces and Yavi in general.

Respondents

Respondents were recruited from the student population in a school of information and library science at a large southeastern research university. All respondents reported extensive experience with both online searching and Internet use. The goal was to have highly experienced searchers assess the usability of Yavi. The bias anticipated from this non-random sample of experienced and educated searchers was intentional. It was hoped that this respondent pool would address the extent to which Yavi is usable to experienced searchers, instead of confounding Yavi's usefulness with general knowledge, or lack of knowledge, about IR.

Cognitive Tests

Two pen-and-paper tests from the Educational Testing Service Kit of Factor-Referenced Cognitive Tests (ETS 1977) were administered: the card rotation test and the map-planning test. Because various researchers have reported a predisposition to prefer or avoid visual interfaces, these tests were intended to aid in discriminating respondents' assessment of Yavi with their visual and navigation skills.

The Tasks

The two tasks were self-administered without supervision. Respondents did not time their performance, but generally reported completion of both tasks in less than one hour. Respondents were able to repeat the training task and refer to the Yavi Quick Reference (Appendix A) as needed. Respondents were able to contact the researcher if any problems or uncertainties were encountered, but none did.

Both tasks involved navigating an Information Space, performing some basic system functions with the space, and answering questions about it. After the second task, respondents were given the opportunity to provide general feedback about Yavi. A combination of numeric and open-ended items were included in the questionnaire.

The first Information Space contained terms from the school's course catalog entries for required courses for MS degrees in information science and library science (here called the COURSE task). This was intended as a familiar subject domain. Each of 12 required courses were placed in the Information Space based the pair-wise correlation matrix of 111 terms from the course catalog. The terms did not include stop terms, faculty names, or terms found in only one course description. A printed copy of the catalog descriptions of the courses was provided for reference. Respondents were given a 2-page questionnaire asking them to identify the closest pair of courses, the closest course to the center of the space, and to assess the appropriateness of various aspects of the space.

The second task was similar, but for an Information Space based on a novel subject (here called the TREC task). A topic from the TREC experiments (Voorhees & Harman, 1997) was chosen, topic 116, which deals with generic drug substitutions (Appendix B). In addition to items asked for the COURSE Information Space, the written questionnaire asked respondents to judge the difficulty of the topic and suggest search terms for the topic (before actually starting Yavi). Respondents retrieved a document using Yavi and judged its relevance to the topic.

After the tasks, respondents were asked the extent to which they believed they understood how to use Yavi, how useful it was, and what sort of tasks Yavi might be useful for. As hoped, the results gave specific insights into further avenues for continued development and evaluation of Yavi. Although the small number of respondents did not produce many results with statistical significance, the open-ended data, when combined with visual inspection of the trends among questionnaire items, give an effective picture of how Yavi is perceived and its potential role for IR.

Limitations and Bias

Limitations to the study have to do mainly with the small sample size and potential bias in the relationship between the researcher and the respondents. Although graduate student assistants were the primary intermediaries between the researcher and the respondents, they were aware of who the researcher is and his faculty status in the school. More favorable evaluation statements might be expected as a result.

Although the number of terms and documents presented by Yavi is not nearly so great as for a real-world information retrieval system, this should not be considered a major shortcoming. Consider that although the number of items is small, these items could have been selected from a much larger dataset. For real-world applications, the anticipated role of Yavi is to visualize and interact with query results (and possibly re-formulate the query). Just as Web search engines, CD-ROM databases and commercial IR systems present search results, not the entire dataset they contain, so can Yavi.

On modern hardware, such as a mid-range consumer PC or workstation, Yavi has been used to present several thousand items with no perceivable delay in any of its functions. We do not anticipate that the relatively small-scale information spaces used for this study indicate any difficulty for Yavi to scale to larger information spaces.

Results

This analysis will focus on describing the trends and characteristics in the data collected, without anticipating whether the data patterns would be maintained in a larger sample. Instead, as stated previously, the emphasis will be on identifying useful areas for further development of Yavi, and in generating ideas about future evaluation studies for visual IR systems.

Can Respondents Judge Distance Accurately?

For each task, respondents were asked to write the closest pair of items, also the item closest to the center of the space. The center was represented by X, Y and Z axes, but there was no tool to measure inter-item distance. The task was further complicated by the 3D view, in which items that appear close might turn out to be further away when the space was rotated (see Figure 1).

Accurate judgment requires respondents to understand that larger items are closer, that rotating the view changes the visual relationship among items, and that distances among items are constant as the view changes. For these judgments, responses may be compared to the (objectively) correct answers.

For the COURSE task, only one respondent correctly identified the closest course to the center, while another two respondents identified one of the other top two courses as closest. The other 13 respondents identified as closest a course that appeared closest before rotating the space (that is, close on the X and Y axes, but not Z). Only one respondent correctly identified the closest pair of courses (not the same individual who correctly identified the closest to the center), while three identified one of the other three closest pairs as closest. All but one of the 11 other respondents was able to identify a pair that appeared close on the X and Y axes.

For the TREC task, 11 respondents correctly identified the closest term to the center, while the other four chose the second closest. Two respondents correctly identified the closest pair of terms, while the other 13 identified one of the other four closest pairs. This drastically different outcome from the COURSE task is almost definitely due to practice, as the COURSE task was completed first. By the time respondents approached the TREC task, they knew more about rotating the space to get different views of term and document relations.

These results indicate that judging inter-pair distances is especially difficult (consider that for 12 courses there are 66 pairs; for the 32 terms in the TREC task there are 496 pairs). There was no ruler or scale, and numeric X, Y and Z locations were not provided. However, respondents were able to identify items close to the center (once they rotated the space), which indicates that basic ability to understand the visualization and make rough-scale judgments about distance/similarity is present. For designers, this result indicates at least two points. The first point is to not assume that numerically different locations will be perceived as different by users. The second point is to provide tools for assessing distance/similarity if such judgments are desirable.

Are the Information Spaces Consistent with Respondent Perceptions?

For each task, an arbitrary set of three pairs of items was chosen (the items were documents for the COURSE task and terms for the TREC task). Respondents were asked to judge the extent to which the displayed distance (similarity) corresponds to their notion of conceptual similarity:

> For each pair of courses below, indicate whether you think the "distance" between the courses is an accurate representation of their conceptual SIMILARITY on a scale of 1 to 5 (circle one):
>
> | 1 | 2 | 3 | 4 | 5 |
> |---|---|---|---|---|
> | Way too far/ different | Somewhat too far/different | About right | Somewhat too near/similar | Much too near/similar |

Responses indicate more variety in assessments of the TREC task than the COURSE task. For COURSE, the grand mean across all 3 course pairs was 3.18, with an overall standard deviation SD=.89. For TREC, the grand mean across all three term pairs was 3.01, SD=1.83. A case-by-case examination of scores supports the interpretation that assessments for both TREC and COURSE centered on "About right," but there was considerably greater variety for COURSE than for TREC. (A note on the use of means: although taking the mean and SD of non-ratio level data is not statistically appropriate, they are in this case a very useful way to characterize the data, and so will be employed throughout this analysis.)

The range of values given also supports the finding of greater agreement for TREC, in that the COURSE assessments ranged from 1 to 5, but TREC ranged from only 2 to 4 across each set of three pairs. A likely interpretation of this result may be found in open-ended comments provided by respondents for these pairs of items, in which they were extremely critical of relationships among courses (most of which they had already spent a semester in) – often commenting on detailed aspects of courses such as the content and who teaches them. There were far fewer comments for the TREC task terms. For COURSE, comments included, "after having both classes, I noticed a lot of overlap," and, "the Master's paper and Intro to Computing class are too close."

At least two alternative interpretations are possible, however. First is that the Information Space method used here does a better job at identifying term relations than document relations. Second is that respondent' perceptions of term relations may be less focused than document relations (perhaps because terms are used differently in different contexts).

Are the Information Spaces Perceived as Accurate?

For the TREC and COURSE tasks, respondents were asked to judge the overall accuracy of the presentation:

To what extent do you think this is an accurate visualization of the conceptual similarity among the courses?

1	2	3	4	5
Very inaccurate	Somewhat inaccurate	Mixed	Somewhat accurate	Very accurate

For the COURSE task, the mean was 2.86, SD=.99, with a range from 1 to 4. For the TREC task, the mean was 3.57, SD=.64, with a range from 3 to 5. These results are consistent with the findings of the previous section, but do not help to distinguish among the possible interpretations mentioned there.

Did Respondents think they were able to use Yavi?

After both tasks were completed, respondents were asked to provide a self-assessment of their level of comfort with Yavi and the extent to which they believed Yavi was useful:

To what extent do you think you understand how to use Yavi?

1	2	3	4	5
Very poor understanding	Somewhat poor understanding	Mixed	Somewhat good understanding	Very good understanding

To what extent do you think the Yavi system may be useful?

1	2	3	4	5
Very useless	Somewhat useless	Mixed	Somewhat useful	Very useful

What are some of the things you think Yavi might be useful for?

What suggestions do you have for the improvement of Yavi?

Responses for both numeric items were mixed, but tended towards the negative. For understanding, the mean was 3.00, SD=1.13, range from 1 to 4. For usefulness, the mean was 3.14, SD=0.86, range from 1 to 4. A number of respondents did not feel comfortable with Yavi and worried they were not able to give correct responses: “I’m really not sure if I did this right,” and, “not really sure what this means.”

Open-ended responses addressing usefulness indicate that respondents did understand at least the nature of Yavi, if not the full detail of its function. Comments included: “easy to see relationship between terms,” “good for searching on concepts the user is completely unfamiliar with,” and, “BI tool for the video-game generation.”

Although the level of confidence about understanding Yavi was fairly low, all 15 respondents were in fact able to utilize the system for all its available functions: moving through the space, choosing various options, and selecting terms and viewing documents. The newness of the interface certainly contributed to respondents feeling uncomfortable with its use. An additional possibility, which will be investigated with a later version of Yavi and larger sample, is that some of the discomfort with Yavi stems from familiarity with more traditional interfaces to library OPACs, CDROM databases, and Web search engines.

Do Cognitive Test Scores Matter?

The inclusion of two standardized cognitive tests was inspired by the guidelines for the TREC-7 interactive track (Over, 1998), in which such tests were given to facilitate comparison of findings across different research sites. Another inspiration came from the INQUERY team at TREC-6 (Allan et al., 1997), where a preference for visual interfaces was found among some respondents, especially those under 40 years of age. (Note that although age was not asked of this study's respondents, the approximate age range was 25-40.)

The cognitive tests used were the map planning test and card rotation test. Both employ pen-and-paper, and involve a training phase followed by a strictly timed 3-minute exercise. For map planning, the exercise is to determine which "building" is passed when en route from one location to another in a grid street map. For card rotation, a target drawing is compared to test drawings, and respondents must choose which of the test drawings are the same as the target drawing but rotated. Scoring for each test is simply the number of correct responses given in 3 minutes. These tests were chosen to investigate whether cognitive skills suspected of being related to the ability to use a 3D navigable visual interface actually impact any of the other items being measured.

The mean score for map planning was 11.26, SD=4.04, range from 6 to 19. For card rotation, the mean was 60.2, SD=16.3, range from 25 to 79. Although the scales are different (people could complete more card rotations in 3 minutes than map plans), the relative ranges and SDs are comparable. The Pearson product moment correlation between the two tests was not significant ($r=.32$, $p>.10$), however a visual inspection of the score distribution does indicate an approximately linear positive pattern among scores.

Further examination of relations among other items measured produced very few scores of statistical significance. (Correlation with 5-point scale items is not statistically appropriate because the scores are not ratio level; Chi-square or Fisher exact tests were used in these cases.)

Visual inspection of the data indicates that card rotation test scores may be positively related to the perceived usefulness of Yavi, but more data are needed if statistical significance is to be achieved. No other trends are evident with these two cognitive tests. The obvious interpretation is simply that more data points are needed to identify trends and achieve statistical significance. Another possible interpretation is that the sample was not varied enough – perhaps the sample (or population of students from which they were drawn) has a tendency towards a more limited range of scores on these tests than the general population. Discarding the idea that abilities or perceptions of visual interfaces are influenced by these cognitive tasks would be premature, but this study failed to find any substantial evidence for the notion.

Conclusion

The importance of visual interfaces for human-computer interaction seems beyond question. Yet the role of visualization for IR systems, such as Yavi's visualization of retrieval results, has not been proven. Many researchers have identified means for visualizing the relations among queries, terms and documents. In most cases, the specific methods were not suitable for large-scale IR systems, or it took a long time for the visualizations to be created.

The Yavi prototype has been presented as a possible approach to the challenge of visualization for IR. Rather than trying to visualize the entire contents of an IR system, Yavi is more suitable for visualizing several hundred, or perhaps a few thousand, items resulting from a query. By providing means for visualizing term-term, term-document and document-document relations, Yavi may be used both for query redirection (by selecting new terms) and for document selection.

The fundamental research result from this work is that the fifteen respondents were able to use Yavi to visually navigate an Information Space. The extent to which Yavi is actually effective for IR, either as a tool for selecting terms and retrieving documents, as a facilitator for browsing search results, or another purpose, was only assessed indirectly through responses to open ended questionnaire items. These responses indicated a potential for usefulness that will be investigated during continued development of Yavi.

Specific points for development, as suggested by the research, include:

- Limiting rotation, especially about the Y axis, to lessen disorientation;
- Providing visual reference points (backgrounds, terrain, etc.) to help enhance depth perception;
- Enabling viewing of relationships beyond similarity, e.g., by selecting a document and seeing all terms in it;
- De-emphasizing keystroke commands in favor of on-screen heads-up menus or pull-down menus;
- Zooming and expanding/exploding the space was available, but underutilized. Train to use these features.

Separating Yavi from its current retrieval system back end is straightforward. We would like to pursue Yavi as a general-purpose visual interface for a wider variety of datasets, notably the outcomes of Web search engine queries. In the Web context, Yavi could enable visualization of search engine results incorporating rich queries and hundreds of candidate retrieval documents – rather than simple queries and short lists of documents most frequently seen today.

Basic concepts for multidimensional similarity-based Information Space need to be communicated to users if they are to be effective. Although some respondents easily understood the basis for term and document locations, most did not. Just as new users of IR systems must learn about Boolean sets and other concepts, Yavi represents a departure from familiar systems and requires some training.

The next phase for implementation and evaluation of Yavi is iterative development (underway) and evaluation with a stronger emphasis on controlled experimental design. The interactive track experiments from the TREC conference have demonstrated the feasibility of conducting comparative research for IR systems and interfaces, but also highlighted the difficulties involved – especially the difficulty of gaining statistical significance even with moderately large sample sizes.

For researchers in IR and HCI, the work here provides substance to the notion that visual, navigable 3D interfaces are usable, at least from a functional point of view. The question of whether 3D is really better than 2D remains open, along with the issue of whether the real problem with 3D is simply the paucity of 3D input and output devices such as those found in virtual reality labs and high end gaming environments. Another open issue is whether visual interfaces for IR can be more effective than text-based interfaces (or what sort of blend is appropriate). With further structured research and with the development and refinement of prototype systems, we can strive to address these issues.

References

Allan, J.; Callan, J.; Croft, W.B.; Ballesteros, L.; Byrd, D.; Swan, R.; Xu, J. 1997. "INQUERY does battle with TREC-6." In Harman, Donna and Voorhees, Ellen (Eds.). Text REtrieval Conference (TREC-6) Proceedings. Gaithersburg, MD: National Institute of Science and Technology.

Borgman, Christine L. 1989. "All users of information retrieval systems are not created equal: An exploration into individual differences." Information Processing and Management 25(3): 237-251.

Card, Stuart K.; Mackinlay, Jock D.; Shneiderman, Ben. 1999. Readings in Information Visualization : Using Vision to Think. San Francisco: Morgan-Kaufman.

Card, Stuart K.; Robertson, George G.; Mackinlay, Jock D. 1991. "The Information Visualizer, an Information Workspace." In Proc. of ACM SIG/CHI '91. New York: ACM Press.

Chalmers, Matthew & Chitson, Paul. 1992. "Bead: Exploration on information visualization." SIGIR Forum (June) 330-337.

Chen, Chaomei. 2000. "Individual differences in a spatial-semantic virtual environment." J. Amer. Soc. for Information Science 51(6): 529-542.

Chen, H., J.; Martinez, A.; Kirchhoff, T. D.; Ng, I.; Schatz, B.R. 1998. "Alleviating Search Uncertainty Through Concept Associations: Automatic Indexing, Co-occurrence Analysis, and Parallel Computing," J. Amer. Soc. for Information Science 49(3): 206-216.

Cochrane, Pauline; Johnson, Eric. "Visual Dewey: DDC in a Hypertextual Browser for the Library User." Advances in Knowledge Organization, Vol.5. Washington DC: International Society for Knowledge Organization, pp.95-106.

Deerwester, S.; Dumais, S.T.; Landauer, T.K.; Furnas, G.W.; Harshman, R.A. 1990. "Indexing by latent semantic analysis." Journal of the Society for Information Science, 41(6), 391-407.

Dillon, Andrew. 2000. "Spatial-semantics: How users derive shape from information space." J. Amer. Soc. for Information Science 51(6): 521-528.

Educational Testing Service (ETS). 1977. Educational Testing Service Kit of Factor-Referenced Cognitive Tests. Princeton: ETS.

Golub, Gene H.; van Loan, Charles F. 1996. Matrix Computations 3rd ed. Baltimore: John Hopkins University Press.

Hawking, David; Craswell, Nick; Thistlewaite, Paul; Harman, Donna. "Results and Challenges in Web Search Evaluation." Proceedings of the Eighth World Wide Web Conference (WWW8). Toronto: Elsevier. pp. 243-252.

Hull, David A. 1999. "The TREC-7 Filtering Track: Description and Analysis." In Voorhees, Ellen & Harman, Donna (Eds.). Proceedings of the 7th Text REtrieval Conference (TREC-7). Gaithersburg, Maryland: NIST.

Ingwersen, P. 1996. "Cognitive Perspectives in Information Retrieval Interaction: Elements of a Cognitive IR Theory." J. Documentation 52(1): 3-50.

Kohonen, T; Kaski, S.; Lagus, K.; Salojärvi, J.; Honkela, J.; Paatero, V.; Saarela, A. (1999). "Self organization of a massive text document collection." In Oja, E. and Kaski, S. (Eds.). Kohonen Maps. Amsterdam: Elsevier. pp. 171-182.

Korfhage, Robert R. (1997). Information Storage and Retrieval. New York: Wiley Computer Publishing.

Kruskal, J.B. & Wish, M. 1978. Multidimensional Scaling. Beverly Hills: Sage Publications.

Mark Pejtersen, Annelise. 1992. "New Model for Multimedia Interfaces to Online Public Access Catalogues." The Electronic Library 10(6): 359-366.

Munzner, Tamara. 1998. "Exploring Large Graphs in 3D Hyperbolic Space." IEEE Computer Graphics and Applications 18(4): 18-23.

Munzner, Tamara; Guimbretière, François; Robertson, George. 1999. "Constellation: A Visualization Tool for Linguistic Queries from MindNet." Proceedings of the 1999 IEEE Symposium on Information Visualization 132-135.

Newby, Gregory B. 1993. "Virtual Reality." In: Williams, Martha E. (Ed.). Annual Review of Information Science and Technology 28. Medford, NJ: Learned Information, pp. 187-229.

Newby, Gregory B. 1997. "Context-Based Statistical Sub-Spaces". Text REtrieval Conference (TREC-6) Proceedings. Gaithersburg, MD: National Institute of Science and Technology.

Over, Paul. 1998. "TREC-7 Interactive Report." In Harman, Donna and Voorhees, Ellen (Eds.). Text REtrieval Conference (TREC-7) Proceedings. Gaithersburg, MD: National Institute of Science and Technology.

Pirolli, Peter & Card, Stuart. 1995. "Information Foraging in Information Access Environments." In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery.

Robertson, George; Mackinlay, Jock; Card, Stuart. 1991. "Animated 3d Visualizations of Hierarchical Information." In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, 189-194. New York: Association for Computing Machinery.

Salton, Gerold; McGill, Michael. 1983. Introduction to Modern Information Retrieval. New York: McGraw Hill.

Schamber, Linda. 1994. "Relevance and Information Behavior." In Williams, Martha E. (Ed.). Annual Review of Information Science and Technology 29: 3-48.

Schamber, Linda; Eisenberg, Michael; Nilan, Michael. 1991. "Towards a Dynamic, Situational Definition of Relevance." Information Processing and Management 26(2): 755-776.

Sebrechts, Marc M.; Vasilakis, Joanna; Miller, Michael S.; Cugini, John V; Laskowski, Sharon J. 1999. "Visualization of search results: A comparative evaluation of text, 2D and 3D interfaces." Proc. of ACM SIG/IR 3-10. New York: Association for Computing Machinery.

Shneiderman, Ben. 1996. "The Eyes Have It: User Interfaces for Information Visuazation." Human Computer Interface Laboratory Technical Report CS-TR-3665. Available: http://www.cs.umd.edu/TRs/groups/HCIL-no-abs.html

Shneiderman, Ben. 1997. Designing the User Interface, 3rd ed. Reading, Massachusetts: Addison-Wesley.

Tufte, E.R. 1990. Envisioning Information. Cheshire, Connecticut: Graphics Press.

Voorhees, Ellen; Harman, Donna. 1997. "Overview of the 6th Text REtrieval Conference (TREC-6)." In Harman, Donna and Voorhees, Ellen (Eds.). Proceedings of the 6th Text REtrieval Conference (TREC-6). Gaitherburg, Maryland: NIST.

Figure 1: Screen shots of Yavi in use for the TREC task.

Figure 1a: Detail of the pop-up Yavi menu. Images in Figures 1a-1d appeared in color during the interface evaluation. Appendix A lists all mouse and keystroke commands.

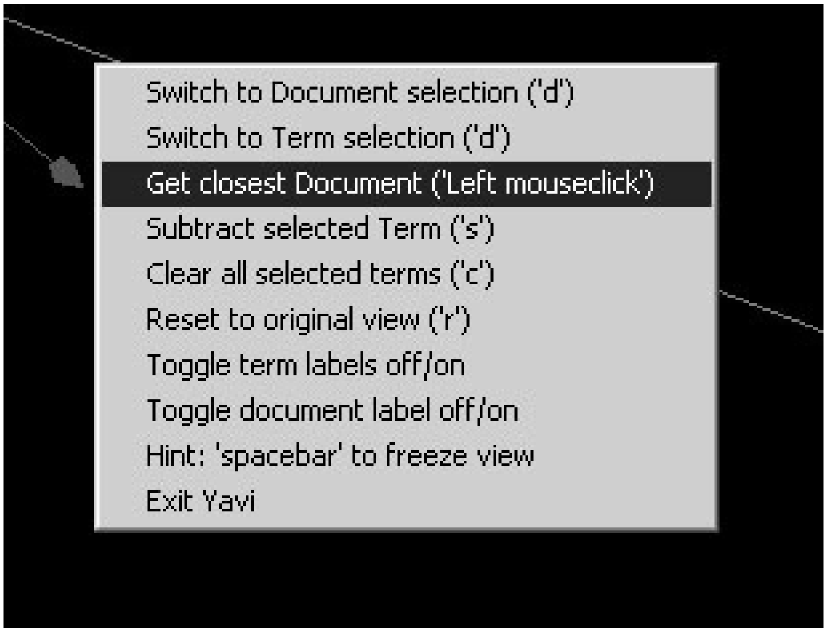


Figure 1b: Detail of Topic 116 term view. Term labels are shown next the each term's location (denoted by a blue point). Document labels are orange points. X-, Y-, and Z-axes are shown to aid in orientation. Moving the mouse rotates the view.

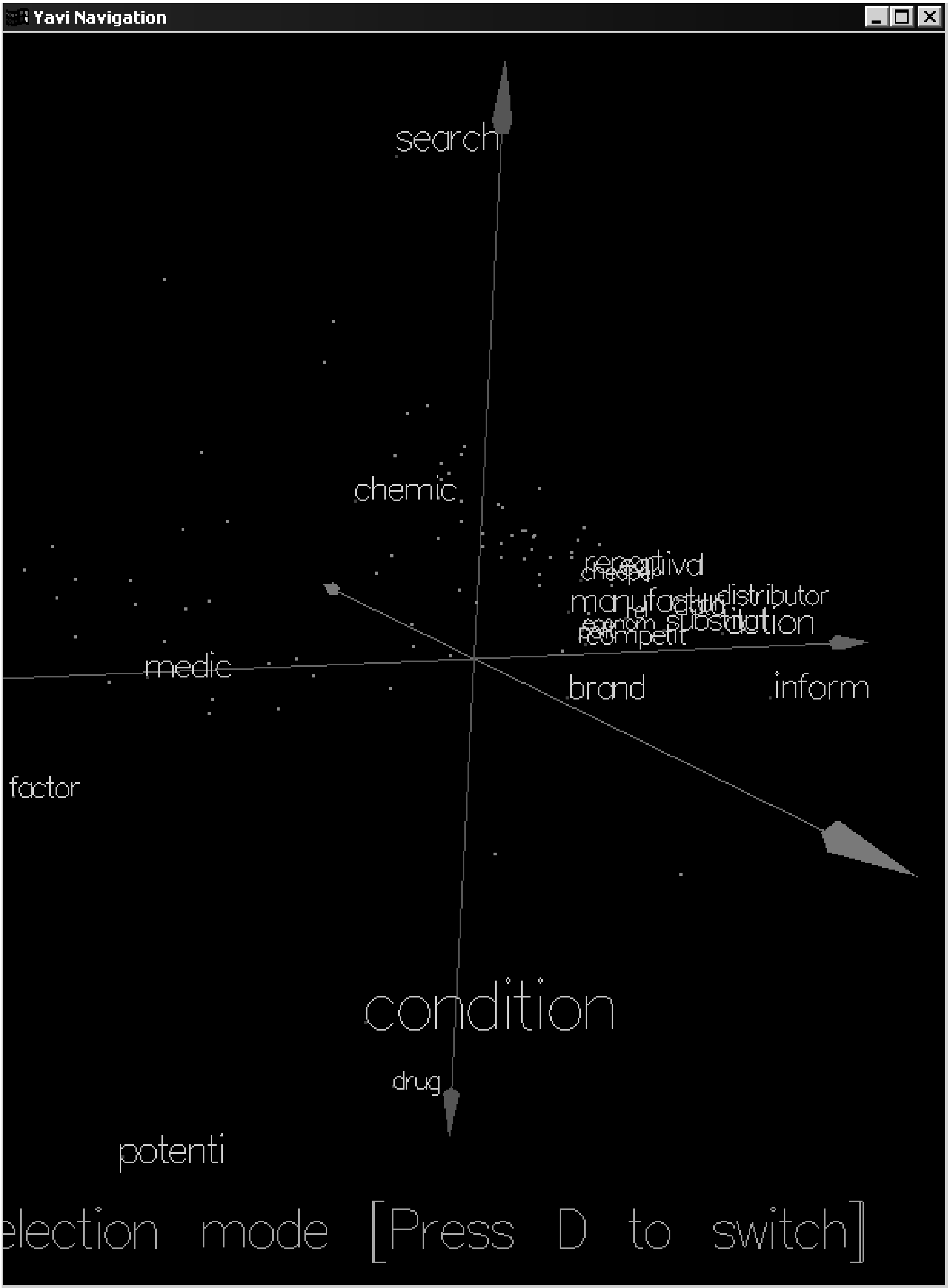


Figure 1c: Detail of Topic 116 with terms "brand" and "manufacturer" selected. Each selected term is given a color, and all documents with that term put up a colored flag. All documents with all selected terms are given a white ball (this is the Boolean intersection of the term sets).

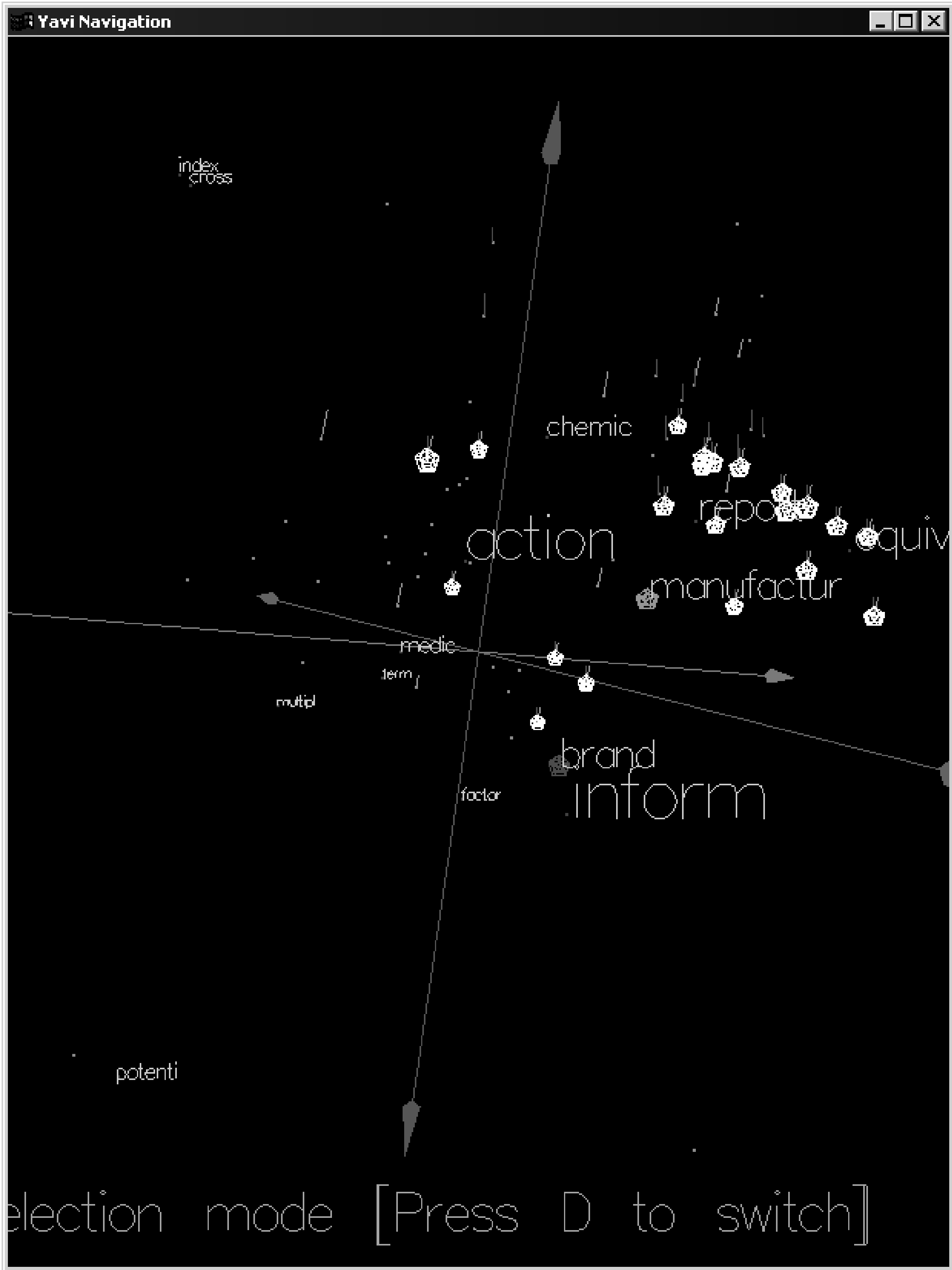
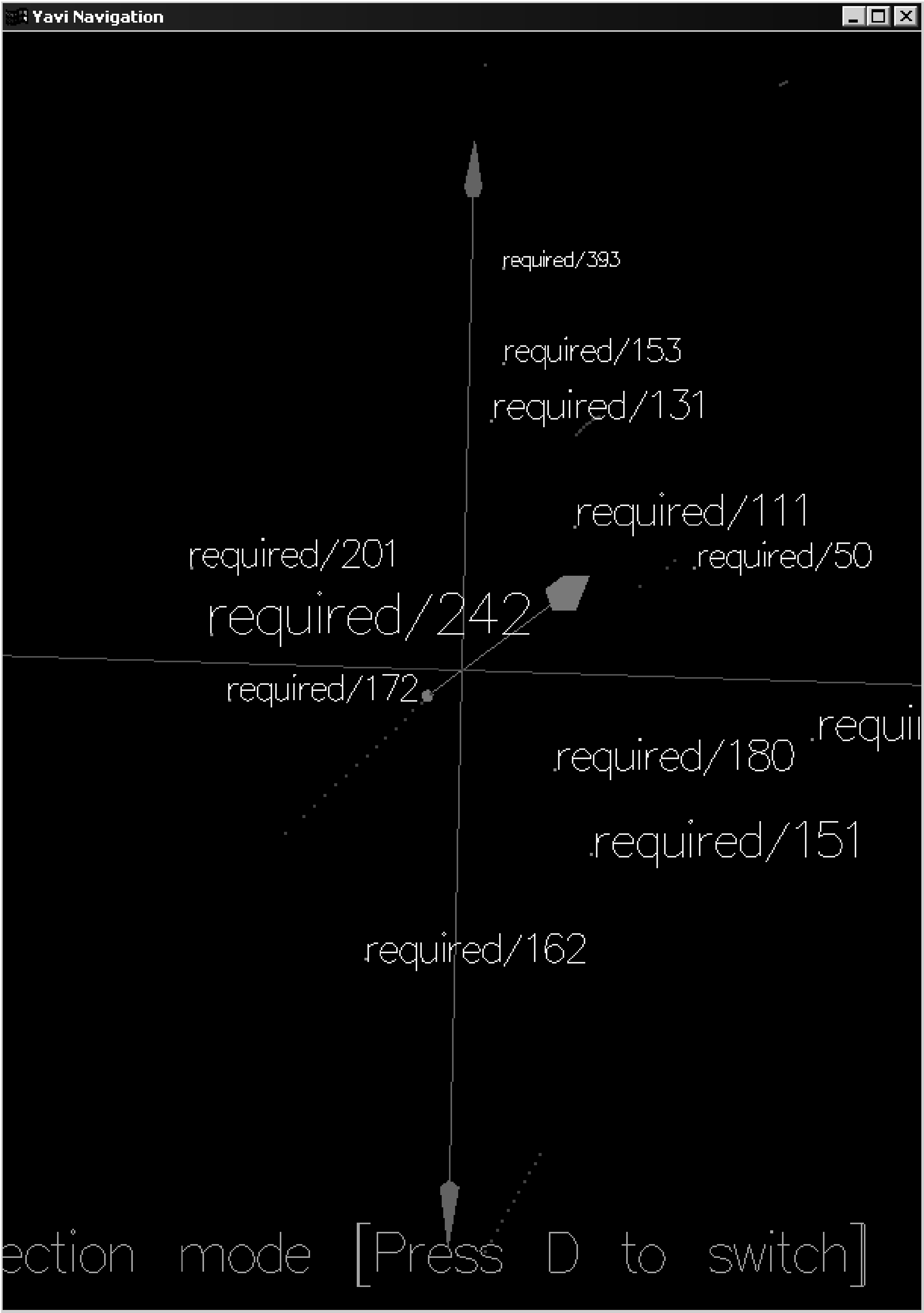


Figure 1d: Detail of Required course documents. Closer documents appear larger. The X-, Y- and Z-axes shown as lines that appeared in red, green and blue.



Appendix A: Yavi Quick Reference

Start Yavi with the name of the Information Space as the first argument. Optionally, a second argument can be a scaling factor (to make things closer or further apart). Samples:

Yavi test1

Yavi qrc .01

The basis: In Yavi, each TERM is positioned based on its tendency to occur with other terms. So, terms that tend to occur together in documents are close together; terms that tend to not occur together are far apart. (Terms that are very close are "nudged" apart, which results in some lines or series of terms.)

DOCUMENTS are located at the center of the terms they contain. This means that documents are not necessarily close to their terms – they might be far apart! But documents that are related should be close to each other.

Modes: Clicking the left mouse button either selects TERMS or DOCUMENTS

TERM mode: clicking the left mouse button ADDs the closest term. The term will be assigned a color, and any documents with the term will put up a "flag" of that color.

- Shift to TERM mode with the "t" key (or use the right-mouse menu)
- Use the "s" key to SUBTRACT the term you ADDed
- Use the "c" key to CLEAR all ADDed terms

DOCUMENT mode: clicking the left mouse button RETRIEVES the closest ACTIVE document, which will be displayed in an Internet Explorer (IE) window. The IE window is independent of the Yavi window - you can close it any time, and you can use the Windows 98 menu bar or ALT+TAB to switch between windows.

- Shift to DOCUMENT mode with the "d" key (or use the right-mouse menu)
- Documents with all ADDed terms will be represented with a white ball. Only these documents can be RETRIEVED with the left mouse button.
- As you ADD more terms, the Boolean set of documents with ALL of the ADDed terms will become smaller, so fewer documents will have a white ball and be retrievable.

Moving the view:

- Moving the mouse will rotate the view, left/right and up/down. Hold down the spacebar to move the mouse without rotating the view.
- Use the arrow keyboard keys to zoom in and out of the space
- Use the plus + and minus - keys to expand and contract the space (this gives more or less separation among the terms and documents in the space).

Other commands:

- Use "r" to RESET the view (in case you accidentally lose sight of the Information Space)
- You can turn DOCUMENT and TERM labels and icons on and off, using the right mouse button.