Emergent Structure in Analytic Workspaces: Design and Use of the Visual Knowledge Builder

Frank M. Shipman III, Haowei Hsieh, Robert Airhart, Preetam Maloor, J. Michael Moore, Divya Shah

Department of Computer Science and Center for the Study of Digital Libraries
Texas A&M University, College Station, TX 77843-3112 USA
shipman@cs.tamu.edu

Abstract: People use visual and spatial cues when sharing information in a workspace, be it physical or virtual. Characteristics of the information and the space, as well as the tools available for manipulation of the information alter the communicative practices that emerge during a collaborative task. The Visual Knowledge Builder (VKB) is a visual information workspace supporting the collection, organization, and annotation of information. VKB emphasizes the emergent nature of information spaces. This emphasis has led to the inclusion of navigable history, a spatial parser for recognizing emergent visual structures, and mechanisms for integrating information from a variety of sources and media types. Use of VKB for note taking, project and conference management, and writing underscores the value of supporting lightweight interpretation in information systems.

Keywords: information workspace, analysis, visual language, emergent structure, history, spatial parser

1 Introduction

When confronted with an information need, many people now go to their Internet connection rather than their television, radio, newspaper, or books. The Internet's navigation and search facilities allow interactive browsing and selection of information. This interactivity, combined with differences in the quality and quantity of information, has led to changes in the practice of information tasks.

One effect of the growth of the Internet is that people have easy access to a greater variety of information than ever before. It is now common to find confusing or contradictory information on many topics. As an example, take the simple task of using AltaVista to search for the "capital of Kentucky." Three of the top four pages are for the city of Berea rather than the capital city of Frankfort. By reading closer we learn Berea is either "the antique capital" or "the folk arts and crafts capital" of Kentucky depending on the source. People now can (and frequently need to) gather corroborating evidence for their information tasks.

The nature of information sources has changed as well – many sources of information are unedited or unmonitored, imposing the responsibility of deciding what to believe and what not to believe on the reader. This is true in print as well but traditional document genres help – flyers on a bulletin board are easily distinguishable from a newspaper, book, or tabloid. Also, most people have fast access to few sources in print and these are easy to distinguish. The ability for individuals with little publishing expertise to create professional-looking information repositories makes distinguishing materials based on visible characteristics more difficult on the Internet.

Information tasks vary greatly in complexity. Finding the current price of a stock, the score of a ballgame, or the location of a conference hotel are short-term tasks where searching for or navigating to the information source dominates the effort leading to the completion of the task. On the other hand, when evaluating the state of a technology or predicting the long-term value of a company, searching for sources of information becomes a much smaller portion of the information task. In these cases, the effort involves deciding what information is needed, what information sources are believable, and how the various pieces of information fit together – in short, analysis.

The growing use of the Internet as an information source and the nature of Internet information sources
lead to a greater need for tools supporting analysis. This paper discusses the design of and experiences with one such tool. The next section characterizes analytic workspaces and describes prior and related work. After this, we describe the Visual Knowledge Builder (VKB) and present observations and lessons from the use of VKB for a variety of tasks.

2 Supporting Analytic Practice

Analysis includes the collection, prioritization, evaluation, and interpretation of information. There are many types of tools that support individual phases of this process. Search engines, portals, and information agents primarily aid the process of locating information for collection and its prioritization through ranking and clustering mechanisms. Similarly, social filtering and ranking can play a role in prioritization and evaluation. But to support the analysis process as a whole requires an environment that includes the ability for users to express their own interpretations. Visual workspaces are our focus among potential environments. Here we describe analytic practice, tools to support this, and prior experience with information workspaces.

2.1 Analytic Practice and Tools

Analysis begins when a person decides that there is an information need and starts considering how to satisfy that need. Early activities include identifying relevant information sources and collecting useful-looking documents through some combination of searching and browsing. DLITE (Cousins et al, 1997) has focused on this issue by integrating direct manipulation query mechanisms into the workspace. Web Forager (Card et al, 1996) and Data Mountain (Robertson et al, 1998) help users interact with a large number of documents through computer-generated visualizations and metaphor. To support this phase of analysis, the workspace must allow the aggregation of information retrieved from various sources and in a variety of representations.

Due to the frequently overwhelming amount of information and the scarcity of time, once a set of documents is retrieved, the analyst browses and prioritizes the documents for further attention. This activity, which has been described as information triage (Marshall, Shipman, 1997), is a rough categorization process where documents are skimmed and clustered into virtual piles or containers indicating the analyst’s initial assessment of perceived value. Our prior experience indicates that as skimming occurs, task-based categories for the documents will start to emerge (Marshall, Shipman, 1997). Whatever affordances the workspace provides will be used in this process – documents will be arranged into lists or piles, assigned visual attributes like color or shape, or put into containers (Marshall, Shipman, 1997).

Outline editors and tree browsers can support such interpretive activities but limit their users to interpretations that can be expressed as a hierarchy. Visual information workspaces are more expressive.

2.2 Information Workspaces

Visual information workspaces have a rich history. Wang’s Freestyle (Levine, Ehrlich, 1995) provided a distributed visual workspace allowing users to collaborate by attaching multimedia messages to documents. While not supporting much graphical coding of documents, Freestyle’s thumbnail view included indications of activity by including visual annotations.

Studies of how people organize information in their physical workspace have been used to inform the design of information workspaces. Malone’s study (1983) of the arrangement of information on desks provides an indication of how different types of work will lead to information spaces with different characteristics. Mander and colleagues (1992) analyzed how people used piles of papers in their office and developed a pile-based interface.

Many visual workspaces have required users to express interpretations in schema-based visual languages using explicit relationships between information objects (Conklin, Begeman 1988; Marshall et al, 1991). A survey of how people arranged information in both physical and virtual spaces led to VIKI (Shipman et al, 1995), the precursor to VKB. This study observed the use of patterns in visual space to represent characteristics of and interrelationships within the information. The expressiveness and ease of use of unconstrained visual expression witnessed in the survey resulted in VIKI’s design emphasis on flexible, and implicit, visual representations (Marshall, Shipman, 1995).

3 The Visual Knowledge Builder

The Visual Knowledge Builder (VKB) is a visual workspace for collecting, organizing, and sharing information following on our prior work on VIKI (Marshall, Shipman, 1995) and the Hyper-Object Substrate (Shipman, McCall, 1999).

3.1 Basic Functionality

The interface of VKB is similar to VIKI. Both systems allow users to create a hierarchy of two-dimensional spaces called collections that can contain other collections or visual symbols. Symbols
are the visual representation of information objects. The placement of symbols into collections is similar to the placement of files into folders in the Windows/Macintosh operating systems.

Users modify visual attributes for collections and symbols to indicate their interpretation of the information. Figure 1 shows a workspace with seven visible collections containing a number of symbols from an undergraduate’s use of the system while writing a paper. The six green collections on the left – labeled Phidias, Storyspace, Doorway, Trellis, IRIS, and Hypercard – contain information about four projects relevant to the student’s task while the fifth collection, colored yellow and labeled Stuff, contains other potentially useful information.

Expression of relationships and categories occurs by placing symbols near one another or placing symbols in a collection. Three of the six green project collections have a thick-bordered orange symbol at the top with a URL pointing to further information on the project. Five of the collections contain brown symbols with chunks of text. These are either the student’s summarization of information found or text taken from Web pages and papers about the projects.

**Navigating** into a collection causes the collection to fill the workspace window, similar to Boxer (diSessa, Abelson, 1986) and VIKI. This exposes more of the lower-level workspace. Figure 2 shows the effect of navigating into the “Stuff” collection in Figure 1.

Most of the symbols in the "Stuff" collection have thick blue borders with white backgrounds. The content of these symbols includes text and URLs pointing to further sources. Examining the symbols in Figures 1 and 2, we see three visual types: thick-bordered orange symbols containing a URL, brown symbols with chunks of text, and white symbols with blue border containing text and URLs.

### 3.2 Visual Expression and Manipulation

As described above, the main mode of interacting with information in VKB is the visual manipulation of symbols and collections in the workspace. These include position, size, border width, font style, transparency, as well as colors for the background, font, and border of symbols and collections. Color was crucial to expression in VIKI. VKB extends the color system by including 10 palettes containing a modifiable set of six background and border color pairs. Additionally, users may "break" the pairwise connection between background and border color pairs. Additionally, users may "break" the pairwise connection between background and border colors and select colors from a full RGB color interface.

Part of expression using VKB is the construction of spatial structures indicating relationships between the symbols. A survey of workspace use in Aquanet and other computational and non-computational settings found people created piles, lists, and repeating patterns of visual symbols to indicate relationships between documents or chunks of information.

![Figure 1: Workspace of student using VKB to write a report.](image)

![Figure 2: Result of navigating into the "Stuff" collection from Figure 1.](image)
(Shipman et al., 1995). VKB includes and improves upon the spatial parser built to recognize these structures and the hierarchic click-selection mechanism for iterating through levels of inferred structure.

The spatial parser recognizes three main types of structures: stacks, lists, and composites. Stacks are overlapping symbols of the same type. Lists are vertically or horizontally aligned symbols of the same type. Composites are repeating visual patterns of symbols of specific types, such as a "list label" above a "list element". All objects in Aquanet had a user-defined type and VIKI's recognition of composite structures required explicit object types. We found that users left many objects untyped. VKB infers types for untyped objects based on visual appearance for this reason.

Figure 3: Three steps in selection of recognized structure.

Figure 3 shows four untyped objects from Figure 2 forming a vertical list of composites. The structure recognized by VKB is indicated by the expanding selections for three sequential mouse clicks on the second symbol from the top. Without implicit types, this would be a uniform list and the second click would select all four objects instead of the top pair.

To determine when untyped symbols are considered to be of the same type, the parser uses a visual difference function that compares the visual properties of two symbols (e.g. their width, height, border width, and border and background colors.) When the value of this difference function is greater than a preset value, two symbols are considered visually dissimilar. To make type an associative property, a symbol only has to be visually similar to one member of an implicit type to be considered a member of that type. While this has the potential to consider very visually dissimilar symbols to be of the same type, this has not been an issue in practice.

Color tends to dominate people's visual categorization of symbols and thus plays a large role in assigning implicit types. Experience indicates that background color is still the dominant factor for an object's type with the border color being used as a modifier. Thus the implicit typing algorithm puts greater weight on differences in background color than differences in border color.

3.3 Embedded History
The visual languages used to express emergent interpretations evolve over time. Because of this, symbols may be given particular visual features early in a task, while later the meaning of those features has changed.

Since people rarely go back to make the information space completely consistent, this can lead to the inability to interpret or misinterpretations later in the task. Also, when more than one person works on a workspace or when long periods of time pass between uses of a workspace, memory of what a particular visual feature means fades.

For these reasons, VKB includes an embedded history mechanism similar to Reeves' work (1993) on INDY and Hayashi's (1998) temporally-threaded workspaces. Users may rewind, replay, or step through the process leading to the current state.

Figure 4 shows a project workspace in VKB. In this figure, the history toolbar is shown below the main visual attributes toolbar. The buttons on the left act like a VCR for playing through the history. The slider in the middle shows where the displayed state is in the event list and allows the user to quickly move to specific states in the history. On the right is the timestamp for the event that resulted in the workspace displayed. Additionally, the user can access a list of the sessions, or clusters of activity.

Returning to a prior state and replaying the development of a workspace provides a quick review of prior efforts. As described in the experiences section, we have seen group project spaces replayed in meetings to remind participants of prior discussions and tasks.

Besides being able to move serially through the history, users may want to return to specific events. Reeves' (1993) network designers found that later
design efforts made understanding earlier comments more difficult but that returning to the original context helped disambiguate the comments. Similarly, changes in the visual coding scheme in the workspace can make interpreting the workspace difficult without the ability to return to the context of prior events. The primary interpretive events in VKB are creating a symbol and object, changing a symbol's visual appearance, editing the content of an object, and moving a symbol into a collection. As a result, the user may request to return to any of these events for any symbol in the workspace.

While history mechanisms have been included in other workspaces, the multiple indices and access mechanisms are crucial to supporting the variety of information needs that arise during analysis.

4 Experience with Use

The Visual Knowledge Builder has been in use for over a year for a variety of tasks. These tasks include taking notes, gathering information, writing, project management, and conference organization.

4.1 Information Gathering

A chemistry teacher from a local high school used the Visual Knowledge Builder during her two-week visit to our campus. Her use of the system was to collect information for her high-school students as resources for their chemistry projects. This effort made heavy use of the ability of symbols to point to information available on the Internet. One way of viewing her use was the creation of a shared visual bookmark list for her students, similar to the use of Eastgate's Web Squirrel (Bernstein, 1996) but with additional annotation.

One of the difficulties the teacher experienced occurred when she wanted part of a Web page, but not all of it. Rather than making a link to the page in the workspace, she took chunks of text from the page and copied them into the system. Initially she had to locate the information, create an object, select and copy the information and paste it into the object. To facilitate this style of including information from other applications, VKB can now create objects with the current contents of the copy/cut buffer.

4.2 Writing

The most extensive use of the Visual Knowledge Builder has been by an undergraduate student from the mathematics department who was investigating different technologies and their potential use. His projects were to write reports about "web workspaces" and "information extraction."

He was given a demonstration of VKB and told to collect information in the system while writing his reports. Each of the two tasks took a month or longer. This time was used to learn the basics on the topic, collect and read information (almost solely from the Internet), and write the report. After looking at the resulting workspaces, we interviewed the student to better understand what was in the workspace, how it was organized, and why.

As the version of the system used by the student did not contain the embedded history mechanism, we collected intermediate states of the workspace: 17 for the "web workspaces" task and 10 for the "information extraction" task. Figures 1 and 2 are the information workspace from early in the information collection process for the first task (the third of the 17 states.) At this point, the student has created collections for information about four projects deemed relevant (PHIDIAS, Storyspace, IRIS, and Trellis) and another collection for "Stuff", which he described in the interview as information he "hadn't gotten to" or that "didn't fit on the left."

At the end of the writing task, the system- and project-oriented collections had all been moved into another collection (called "Hypermedia Applications") and the information visible from the top-level space was content written by the student or very general information like definitions. This change in the character of the task from the collection of source materials to writing was echoed in the interview when he said that by then he "understood more" about the domain and the top-level collection now contained information for his main task, to "write the report."

![Image](Image)

Figure 5: Visual coding of information during writing.

The development of visual types and the existence of patterns within the layout of information
occurred early in the writing task. These patterns evolved during the writing process until, near the end of the task, almost all information was included in the visual patterns of information. Figure 5 shows the information that the student decided not to use in his report. The seven visible (green) collections containing information on specific systems follow a pattern of an evaluation of relevance (red object) and a URL (in orange) above chunks of text about the project (brown). While some collections do not contain all of the pattern’s elements, the pattern is clearly recognizable.

The second writing task, a report on “information extraction,” resulted in an information space quite different from the first task. The top-level collection contained partially coded symbols and collections in the top half of the space but uniform and frequently overlapping symbols below. In contrast to the first task, source documents and student writings shared the top-level space.

Comparing the layout of information during the two tasks, there is an apparent difference in the amount of regular structure. Early in the first task, there was a visual coding scheme for source materials and student notes. The regularity of structure was much lower in the second task. Despite the visual clutter, the student readily explained the use of color for grouping information into general concept and project specific collections, with white indicating text the student had written and blue being links to Web-based information.

Looking back at prior experience indicates one possible explanation for the difference between the two tasks. In a study comparing the use of VIKI to the use of paper technologies for a short-term analysis task (Marshall, Shipman, 1995), it was noted that users of the system were less concerned with reading documents and more concerned with the comprehensibility of their results (improving the visual structure.) Could the initial emphasis on visual regularity be a result of the novelty of the visual workspace?

When asked directly about the regularity of the first space and the lack of regularity in the second, the student said there “are many different approaches” to organizing the information in the second task and the categories were still in flux. The student’s recognition of alternative structures led him to postpone adding structure until he understood more. He also indicated he was more interested in the second topic. These comments indicate that the perceived complexity of the domain, perhaps influenced by enthusiasm, led to differences between the two spaces.

Of the uses of VKB, the writing tasks best match the analysis task described at the start of this paper. The experience shows the evolution of the workspace as the user's emphasis changes from the collection of materials to the organization and interpretation of these materials while writing.

### 4.3 Project Management

VKB has been used to help manage three research projects at Texas A&M University. The projects include one or two faculty members and between two and five graduate and undergraduate students involved in day-to-day activities. The research groups use VKB to record current efforts, potential future tasks, and issues and concepts that arise during group discussions. The project information spaces are accessible by all project members although almost all editing activity is performed in the meeting in front of all the project members.

Project meetings take place in a conference room with the VKB workspace projected onto a five-foot diagonal touchscreen. The project's information space becomes a partial record of the meeting. As new activities are brought up -- such as potential system features, evaluations, or writing tasks -- they are added to the workspace for later consideration. Additionally, the information space acts as a "to do" list, provoking discussion of tasks considered important or having approaching deadlines.

Figure 4 shows the Walden's Paths project space. There are two major collections: one containing the project's "To Do List" and the other listing current research directions and potential audiences for papers on that research. This particular figure shows the workspace after five months of use.

The to-do collection is divided into "Real Soon", "Soon" and "Sometime" lists. New tasks are added to lists based on their priority and are moved from one list to another as that priority changes. In this space, the border width of an item indicates whether the task is currently under development, with the thickest borders being for tasks almost completed. Once completed, the tasks are moved to a collection not shown in the figure that contains prior tasks.

The VKB workspaces for the other projects are similar, although each group developed its own visual language for representing information about the project. With two of the project spaces now including more than a year of notes from weekly meetings, the long-term emergence of visual structure in these workspaces can now be described.

As a workspace ages, the increasing variety of information needs for which it is used results in greater complexity of the information in the
workspace. This tends to force the visual language of the space to become more expressive. For example, project spaces initially used pairs of border and background colors to represent one characteristic, such as whether a task involved system building, evaluation, or writing. Later, border and background color came to represent different characteristics of the task. Similarly, only two values of border width were initially used to represent project activity. From this start, a wider variety of border widths were later used to indicate both current and cumulative activity on a task.

The visual workspaces have become central information resources for the projects. As the project members change or forget prior discussions, the workspaces can be consulted to fill in gaps in a member's knowledge. The navigable history allows easy access to the creation and modification dates of each task, concept, and note. By replaying the history of the space, members of the group can witness the identification of tasks, changes in task priority and effort, and task completion or removal over time. Another use of the history that we see is to determine the length of time an idea has been worked on (or not worked on.) For example, meetings occasionally include going back to the creation date of a task to determine the date when a topic was brought up or put on the "to do" list. In short, the workspace history becomes a record of the project's intellectual development.

Experiences with the use of VKB for project management are our best examples of the long-term emergence of visual structures and the variety of ways embedded history changes the role of information workspaces.

4.4 Conference Organization

Another use of VKB has been the organization of the ACM Hypertext 2000 Conference program. This activity included both individual activity by the program chair (first author of this paper) and collaborative activity by projecting the workspace at the program committee (PC) meeting.

VKB was initially used to focus discussion and provide a visual sense of effort on the second day of the PC meeting. At that time, the workspace contained lists of the accepted and undecided papers. Activity at the meeting was centered on these lists: (1) each undecided paper was discussed and placed in either the “accept” or “reject” collection, and (2) the accepted papers were categorized and placed into sessions. Figure 6 shows the arrangement of accepted papers into sessions a month after the PC meeting. Border color indicates the area of work while the height of the symbol distinguishes between long and short papers.

But why use VKB? One reason is the ease of recording the PC's decisions by changes in the workspace. Moving or coloring the symbol for a paper represented to all PC members the effect of the decision just reached. Also, the workspace allowed activities to be flexibly interleaved by moving in and out of collections. This was important because the acceptance of previously undecided papers influenced the topics for categorization and development of the sessions. Likewise, the categorization and grouping of accepted papers into sessions provided information about the overall program that was used in making further acceptance decisions. Also, if there was a question about an earlier decision during the meeting, the history of the decision’s impact on the space was available.

Figure 6: Accepted papers being placed into sessions.

After the meeting, the VKB space was further used to plan the timetable for the 2 track conference. This space used references to the sessions shown in Figure 6, along with similarly color-coded objects for panels, keynotes, and other venues, to more easily create a schedule that avoided similar topics being presented in simultaneous sessions. The visual representation enabled the rapid design of a schedule where this constraint was rarely broken.

Conference organization in VKB shows how a hierarchy of workspaces enables moving between interrelated tasks. It also exemplifies the use of visual cues for collaborative categorization and for design constraint satisfaction.

5 Conclusions

The growing use of unedited information sources increases the need for people to become information analysts, looking for corroborating evidence, deciding what sources are believable and developing interpretations based on information from numerous
sources. Analytic workspaces enabling human-generated visualizations provide a place for this activity to occur.

The Visual Knowledge Builder is designed based on experiences with prior analytic workspaces. Users create a hierarchy of two-dimensional spaces for collecting and arranging visual symbols containing information. VKB:

- enhances the spatial representation by focusing on the visual characteristics found most useful in prior work,
- extends the structures recognized by the spatial parser by identifying object types using visual features,
- increases the variety of external information that may be collected in the workspace, and
- supports comprehension of visual interpretations with an embedded history mechanism, letting users view the emergence of the visual language in a workspace or return to a particular interpretive event.

Use of VKB has included note taking, writing, project management, and conference organization. The emergence and evolution of visual representations for information in these tasks underlines the advantages of flexible visual representations for collecting and sharing information and the need for mechanisms to aid in the generation and interpretation of structure in shared workspaces.

This research emphasizes the human-interpretation of information through emergent visual languages. Current development on VKB combines the authored visualizations of unstructured information described in this paper with computer-generated visualizations based on recognized structures or domain-specific information. Such a combination will allow future analytic workspaces to better support all phases of the analysis process.

Acknowledgements
This work was supported by grant IIS 9734167 from the National Science Foundation and by a grant from FX Palo Alto Laboratory. We thank John Leggett, Cathy Marshall, and Susana Tosca for their helpful comments on drafts of this paper.

References
Malone, T. (1983), How do people organize their desks? Implications for the design of office information systems. ACM Transactions on Office Information Systems, 1, 1, 99-112
Reeves, B. (1993), Supporting collaborative design by embedded communication and history in design artifacts. Ph.D. Dissertation, Univ. of Colorado.