Journal of Universal Computer Science, vol. 7, no. 6 (2001), 517-529
submitted: 1/5/01, accepted: 1/6/01, appeared: 28/6/01 © Springer Pub. Co.

Discovering Knowledge Through Visual Analysis

Jim Thomas
(Pacific Northwest National Laboratory
Richland, WA 99352 USA
Jim.Thomas@pnl.gov)

Paula Cowley
(Pacific Northwest National Laboratory
Richland, WA 99352 USA
paula.cowley@pnl.gov)

Olga Kuchar
(Pacific Northwest National Laboratory
Richland, WA 99352 USA
Olga.Kuchar@pnl.gov)

Lucy Nowell
(Pacific Northwest National Laboratory
Richland, WA 99352 USA
Lucy.Nowell@pnl.gov)

Judi Thomson
(Pacific Northwest National Laboratory
Richland, WA 99352 USA
Judi.Thomson@pnl.gov)

Pak Chung Wong
(Pacific Northwest National Laboratory
Richland, WA 99352 USA
Pak.Wong@pnl.gov)

Abstract: This paper describes our vision for the near future in digital content analysis as it relates to the creation, verification, and presentation of knowledge. We focus on how visualization enables humans to make discoveries and gain knowledge. Visualization, in this context, is not just the picture representing the data but also a two-way interaction between humans and their information resources for the purposes of knowledge discovery, verification, and the sharing of knowledge with others. We present visual interaction and analysis examples to demonstrate how one current visualization tool analyzes large, diverse collections of text. This is followed by lessons learned and the presentation of a core concept for a new human information discourse.

Keywords: Information visualization, knowledge management, digital content and media, digital libraries, visual paradigms, higher-order interaction, human-computer interaction
Categories: H.5, H.5.1, H.5.2, I.3, I.3.8,

1 Visualization, Learning, Knowledge and Communication

In the emerging information age, a critical issue is how we learn from the huge amount of information that bombards us every day through every aspect of life. Recently, researchers at the University of California - Berkeley [Lyman & Varian 2000] reported that the world produces one to two exabytes of unique information every year. That is one billion gigabytes (10^{18} bytes) of text, numbers, images, sounds, and other forms of information that are deemed important by humans for different purposes.

Information visualization is one solution to this vast problem of information overload. This emerging field endeavors to create visual representations of abstract information such as text documents, images and videos, hierarchical and network graphs, and all kinds of information available via the World Wide Web [Card et al. 1999; Tufte 1983, 1990, and 1997; Gershon & Eick 1997; Ware 2000]. With roots in scientific visualization [McCormick et al, 1987], application of visualization techniques to data mining and knowledge discovery tasks began almost immediately [Fairchild 1988].

The power of visualization lies in its ability to convey information at the high bandwidth of the human perceptual system, facilitating recognition of patterns in the information space and supporting navigation in large collections. Text visualization systems offer a variety of approaches to presenting information about collections of text, from conceptual maps [Lin 1992; http://www.pnl.gov/infoviz/] to tools that base their layout on metadata [Nowell et al. 1996; Ahlberg & Shneiderman 1994] or similarity to query terms [Olsen et al. 1993; Spoerri 1993; Hemmje et al. 1994]. Other systems show query term occurrence within individual documents [Hearst 1995], the conceptual structure of individual documents [Miller et al. 1998], thematic trends over time within a collection [Havre et al. 2000], and so forth. Such visualization systems can provide significant value for exploration and insight in text collections. Details of a sample analysis that illustrates this point are provided in section 2.

Expertise and techniques in visualization, statistics, and cognitive science for visualizing large amounts of data have been applied to the emerging discipline of Knowledge Discovery and Data mining (KDD) to form the study of Visual Data Mining [Keim & Kriegel 1996; Rbarsky et al.1999; Wong 1999]. This new approach integrates the human mind's exploration abilities with the enormous processing power of computers to form a powerful knowledge discovery environment. The technology builds on visual and analytical processes developed in various disciplines, including scientific visualization, data mining, statistics, and machine learning with custom extensions that handle very large, multi-dimensional, multi-variate datasets. The methodology is based on both functionality that characterizes structures and displays data, and human capabilities that perceive patterns, exceptions, trends, and relationships.

In the field of information visualization, high expectations have surfaced as users become more familiar with what is possible and more demanding about what they require. Users are no longer satisfied with a single visualization that provides a constant view of unchanging data. The challenge for builders of interactive systems is to create an environment for discovery, verification, and knowledge sharing between systems and people to allow each the ability to learn and adapt from the experience.

Meeting this challenge requires advances in some technologies and consolidation of others. For instance, the more advanced interactive systems must be capable of presenting information to users as well as collecting information from them. These systems must understand the preferences and requirements of individual users, and they must use this information to realize when users require assistance making inferences, when it is appropriate to make generalizations about tasks or information, and when significant events occur that should induce updates to the system's knowledge base. All of these advanced features require that the system can perform complex reasoning about the information it has recorded. Computer applications that do this sort of reasoning exist but typically for closed domains and rigid representation structures [Rich & Knight 1991; Russell & Norvig 1995]. The flexibility, extensibility, and individualization required by more advanced information visualization applications dictate that more universal approaches must be found. This requires that the computing environments are able to represent, and reason about, the information required for effective communication, domain specific information, and the various relationships among this data.

The challenge is not only one of knowledge representation but also one of context representation. Visualization environments of the future must have a deep understanding of communication, and of the users, to effectively assist users make insightful observations and arrive at new and interesting conclusions. Key is the ability to represent various levels of knowledge and context — both that of the system and that of the user — and then communicate this knowledge through an interactive dialogue engaging the human visual system. Our emphasis is on visual communication because we can use the human visual system to take advantage of the high bandwidth between information and the brain.

In the next section, we discuss one major tool that has evolved from our information visualization research at PNNL. This research focuses on dealing with large volumes of textual information. In the course of this work, we have learned much about how users can gain knowledge from large bodies of textual information, how they can formulate and test hypotheses about the "story" behind the information, and how they can use visualization to communicate to others.

Experience with users of many visual information analysis tools has led us to our vision for a new approach to interacting with information — a new human information discourse. This vision is described in section 4.

2 Visual Information Analysis

In this section, we present a fictional user named Mary to illustrate visual information analysis and the knowledge discovery process. Mary uses SPIRE to explore a collection of news stories from the week following the April 1995 bombing of the Federal Building in Oklahoma City. The data Mary fed into SPIRE was plain text with tagged fields for the source location and the date of the story that Mary defined for SPIRE.

SPIRE's text engine applies advanced statistical methods to identify the key topics within a document set — the words that best discriminate among the documents — and produces a document vector, or numerical representation, of each document's essence as it relates to other documents in the set. The document vectors are used by a projection algorithm to produce a two-dimensional numerical representation that can be plotted on a computer screen, creating the ThemeView^{TM} and Galaxies

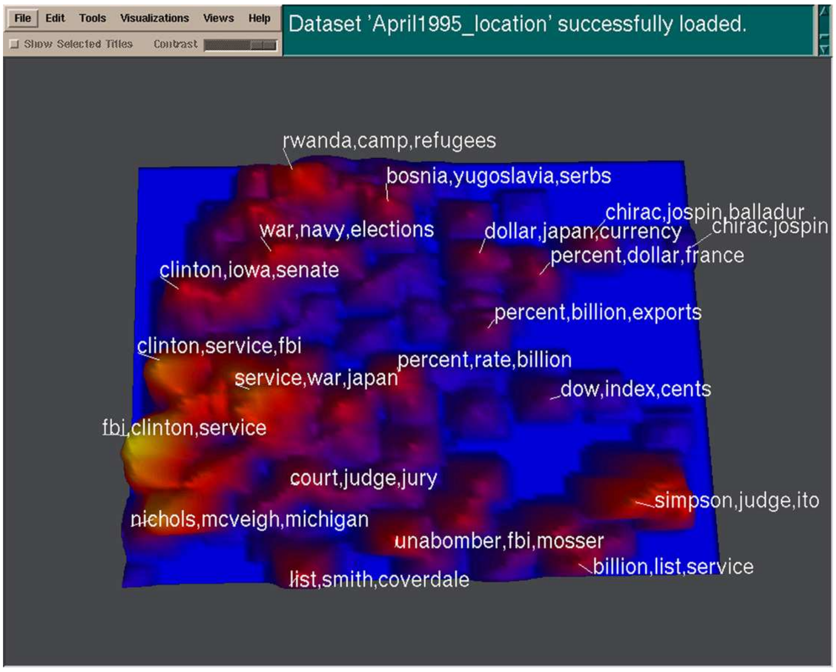

Figure 1: SPIRE ThemeView™ Visualization

visualizations shown in Figures 1 and 2.

Mary chooses ThemeView to begin her analysis. The ThemeView visualization is commonly used as a starting point for exploratory analysis of a collection, because it provides a quick overview of thematic content and orientation to SPIRE's spatial layout of those topics. The ThemeView shows the conceptual content of the collection as a topical landscape, in which hills and mountains signify concentrations of content — frequent mention of closely related words. The more significant the concept in relation to the collection, the higher the peak that represents it. In SPIRE visualizations, proximity denotes similarity. That is, concepts that are closely related in some way are closer together, while those that are different are more widely separated in the visual space.

In this collection, the theme represented most strongly is represented by a high, light gray peak in the lower left corner labeled "fbi, clinton, service" in Figure 1. Mary notices that other nearby high peaks with light tips bear the labels "clinton, service, fbi" and "nichols, mcveigh, michigan." These labels all relate to some aspect of the bombing, investigation, and suspects, as might be expected. In the lower right corner she sees a high peak labeled "simpson, judge, ito," reflecting the fact that the murder trial of O.J. Simpson was in progress at the time of the bombing in Oklahoma City. The top half of the ThemeView has a variety of lesser peaks that represent contemporaneous world events, ranging from violence in Rwanda to fluctuating currency values and political elections. Note that words appearing together as a peak label are not a phrase recognized by the text engine; they are simply themes that are both evident at that point in the collection, though not necessarily in the same documents.

Next, Mary uses the SPIRE Galaxies visualization for the same data set, as shown in Figure 2. This visualization allows Mary to quickly identify concentrations of documents by thematic content. It uses the same projection, or conceptual map, that underlies the ThemeView visualization, so Mary is able to maintain her orientation in the information space. Each point or dot in the Galaxies visualization represents the text for one news story. The distance between points indicates their thematic similarity. Thus, if points in the visualization are close together, then it is likely that the corresponding documents will contain conceptually similar information. If they are far apart, the documents probably will be conceptually diverse. (Note: We have observed that the spatial layout of the points is not inherently meaningful to users, who learn that the spatial relationship among the points indicates conceptual relationships.)

SPIRE's text analysis engine clusters documents using either of two standard algorithms: hierarchical clustering or k-means clustering [Rasmussen 1992; Jain & Dubes 1988]. The larger open circles in Figure 2 show the location of cluster centroids, and the associated text lists the themes that occur most often in the documents for each centroid. Placement of the centroids, like that of the dots, depends on similarity to other centroids, but, as with the dots representing documents, position is not inherently meaningful to users. In the Galaxies shown in Figure 2, Mary finds many clusters with labels that clearly relate to the ThemeView peak labels. For example, in the lower left quadrant, which the ThemeView showed to be most closely related to the Oklahoma City bombing, are clusters labeled "bombing clinton oklahoma," "bombing oklahoma federal," and "bombing nichols oklahoma." Mary also sees similarities between other cluster labels and the ThemeView peak labels for the same regions, such as that for "simpson judge ito" in the lower right corner.

The Galaxies and ThemeView visualizations are rich sources of insight into the thematic content of the collection. By examining ThemeView labels, Mary quickly becomes familiar with the general topics and themes represented. Rapid insight into collection content encourages Mary to begin asking questions about the collection and relationships among the documents therein. Because the system has provided information about collection content, Mary doesn't waste time seeking information that is not present, and the peak and cluster labels provide her with visual clues about the vocabulary used to represent the documents.

Turning back to the basic Galaxies visualization shown on Figure 2, Mary notices that a Galaxies cluster near the bottom and center is labeled "sacramento california bomb." Mary is puzzled, because she knows the bombing story that dominates the collection centers in Oklahoma City. Looking at the ThemeView in Figure 1, she finds a peak in the same location with the words "unabomber, fbi, mosser." "Unabomber" was the nickname given to a terrorist who was attacking prominent U.S. university professors and corporate executives with mail bombs during this period.

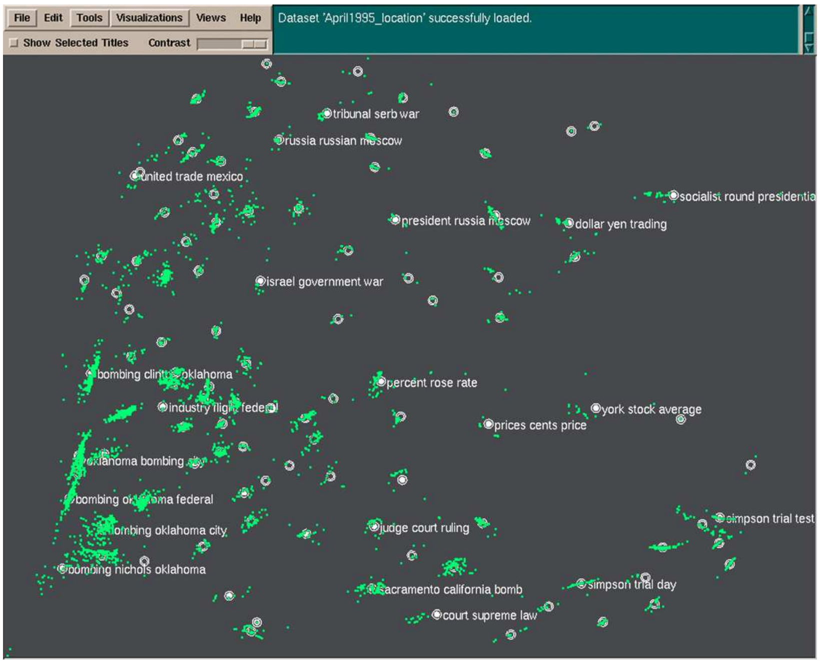

Figure 2: SPIRE Galaxies Visualization

Mary decides to try a few of SPIRE's analytical tools to further explore the collection. She knows that she can turn on or off document titles and cluster centroid labels individually or in groups. She uses the Probe Tool in the ThemeView to see a list of themes ordered from strongest to weakest associated at various locations on the screen. Probing the peak in the lower left corner that is labeled "nichols, mcveigh, michigan," Mary sees themes nichols, mcveigh, fbi, michigan, motel, etc. For the peak labeled "unabomber, fbi, mosser," Probe shows Mary the themes unabomber, fbi, mosser, murray, list, letter, service, etc. And she find that the "simpson, judge, ito" peak has themes simpson, judge, ito, jurors, deputies, court, jury, etc. Next, she uses the Gisting Tool to see frequently occurring terms in a group of selected documents; the tool also reports the number of documents in the set that contain each word. Selecting and gisting the clusters about the Oklahoma City bombing, Mary finds frequent occurrences of bombing, oklahoma, city, nichols, mcveigh, federal, suspect, fbi, truck, etc. Gisting the cluster about the Unabomber, she finds frequent occurrence of sacramento, california, bomb, unabomber, killed, people, package, san, francisco, bombing, office, federal, Monday, etc.

Mary can also create graphical representations for certain types of fielded data with the Field Marker tool, shown in Figure 3. Mary associates colors with the dates following the bombing and uses lines in different orientations to represent several locations of interest, such as Oklahoma City and Washington, D.C. This analysis shows an initial flood of new stories out of Oklahoma City. Over the next few days, she observes a progression of reports from Washington, D.C. and other cities around the world, as politicians and others react to the bombing.

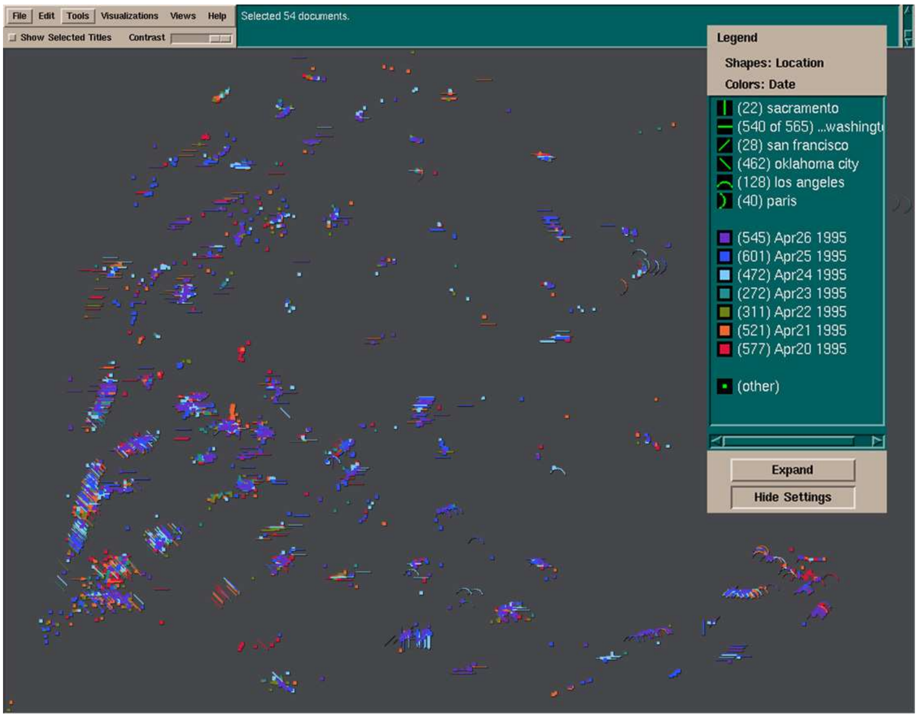

Figure 3: SPIRE's Field Marker Tool

Mary decides to use the TimeSlicer to examine the progressive development of themes. The TimeSlicer lets Mary watch the increase and decrease of various topics in the collection, day by day, as a rising and falling of the ThemeView peaks. The peak for the Unabomber stories first appears as a low, unlabeled peak on the fourth day after the Oklahoma City bombing and rises quickly to be a strong story by the fifth day. Mary wonders if there is some special relationship other than use of bombs that draws Unabomber and Oklahoma City stories together. She decides to explore further, using SPIRE's Query and Group Tool capabilities.

A Boolean "Words in Document" query on "Oklahoma" reveals 1240 news wires containing that word. Another query on "unabomb*" shows that 51 of the stories contain reference to the Unabomber. Mary uses the Group Tool's set operation for Intersection to identify two stories that mention both Oklahoma and Unabomber. Examining these stories, she quickly finds an explanation. The Unabomber mailed three letters to the *New York Times* before striking again. He sent another letter bomb on the day after the Oklahoma City bombing, prompting speculation that this attack was motivated by jealousy over the attention paid to the Oklahoma City bombing.

Recognizing the relationships among these stories and locating the key articles that provide an explanation has taken Mary only a few minutes using SPIRE.

Together, the SPIRE visualizations and analysis tools enable users to quickly perceive the main themes within a collection of documents, locate documents relevant to the topic of interest, and determine where to spend that most valuable resource — human attention. We believe the need for this capability will only increase, given the explosive growth of information.

We have found that visualizations take the mystery out of interacting with a system and the information, allowing the system to present its knowledge of the data for immediate use. Users no longer query blindly, guessing at collection content and keywords without context. Visualizations improve insight and orientation, quickly leading to the better questions that are the real key to discovery.

3 Lessons Learned

Our experience with our clients and visual information analysis tools has yielded some valuable lessons and insights. As we have seen with Mary, people are capable of thinking and interacting with information in many ways that are not supported by traditional user interfaces with their windows, icons, menus, and pointing devices. In particular, people quickly learn to handle rich visual complexity. For example, we all navigate quite well in heavy rush-hour traffic, despite an intense and continuous stream of visual stimuli. We know what is important and focus on that while ignoring the unimportant.

In the course of visual information analysis, measures of what are important really need to come from the analyst. Furthermore, because of the dynamic nature of the analysis process, what is considered to be important shifts as the analysis proceeds and the analyst gains additional insight. SPIRE and other visual analysis tools do not provide this kind of flexibility.

We also know that critical information seldom resides in a single document. Often patterns of relationships among documents are the key to understanding an event or situation, and visualizations support speedy perception of such patterns. We saw an example in the sample analysis, when Mary was curious about the proximity of the Unabomber peak and cluster to those about the Oklahoma City bombing.

We have also learned that increased scale and complexity changes everything. Solutions that work for small collections, from the document vectors to search algorithms to the visualization themselves, strain under the load of large collections. Our clients want the ability to analyze a million documents per day, and doing so requires fundamentally new methods of document analysis and ingest. We also need ways to fuse complex data from multiple sources and multiple media, some of which may arrive in dynamic streams during the analysis.

Finally and perhaps most importantly, we know that people approach information analysis tasks with considerable knowledge and situation-dependent information that they want to bring to bear on the problem at hand. Analysts want to share insights and discoveries with the system, seeing those insights and discoveries reflected in the visualization and sharable with other analysts. They want systems that are responsive to their individual circumstances, to the rich and varied context of collections, and to the unique challenges of rapidly changing situations. Our current systems have no means of capturing that knowledge and situation-dependent information. Furthermore, as analysts work with a collection, they learn and develop insights about the data, the underlying situation, and the problem at hand. In our story, Mary may have concluded that the Unabomber stories, no matter how interesting, were irrelevant to her. She needs a way to tell the system to pay less attention to these items, or to pay more attention to stories that do interest her.

In short, we need ways to capture and integrate what has transpired during the analysis process and what the analyst has learned in the course of the analysis. Some of this can be captured as metadata and associated with the original data set. Some of it becomes new information and knowledge that needs to be fused with what already exists. All of the information must be usable by the analyst and by the system as stories are constructed to communicate their interpretations of the underlying data. We must be able to represent and use the entire context for the communication, including the user's context, the context of the task, and the historical context in addition to the original data.

4 New Human Information Discourse

Our experience has led us to a vision for a new human information discourse that creates a two-way, highly interactive dialogue between the human analyst and the visual information analysis system. This method of discourse will greatly enhance support for learning, discovery, verification, and sharing of knowledge. The new human information discourse is about actively engaging people in conversation about and with information, so the supporting analytical tool learns from its users and actively shares information with them. The visual analysis displayed on the screen and underlying knowledge representations change on the fly in response to user actions, both physical and verbal.

As the information analysis system's way of communicating with its users, a visualization represents the system's view of a collection of information. The human analyst is able to see and interact with the visual representation to learn about the collection and how the system operates. We envision a system that will let analysts use speech and gesture to share information with the system, adding information about the problem at hand, the current state of the world, and the analyst's current hypotheses. In particular, the system will learn from the analyst about which information is more important and why that is so. As the system responds by changing its visual representation, the analyst can see immediate feedback on the system's revised view of the collection and the situation.

Past experience shows that this works best when the user is highly engaged as an active participant. Our vision for a new human information discourse accommodates such users through a much higher-order interaction capability that enables even deeper learning engagements, in context, with many different types of digital media.
Foundations for this new mode of discourse include:

1. User modeling mechanisms for representing and evolving the analyst's knowledge, preferences, and situation.

2. Ways to extract and represent knowledge about the analytical challenge at hand – the characteristics of the problem and information need.

3. Methods of human-system dialog that support resolving ambiguity in questions and user actions, to support the analyst in sharing knowledge with the system.

The information discourse of the future will be based on flexible, interactive conversations between the user and machine. The results of these conversations will be story-like constructs [Schank 1995] that help the user communicate with others, records the analysis process, and helps the user manage multiple hypothesizes and conflicting evidence. Key to this experience will be storytelling via digital media-enhanced communication. All forms of digital media such as text, images, sound, and video will be used with agents aiding the analyst and the system with a variety of routine and time-consuming tasks. This discourse will help bridge the interfaces between the user, the information, and the situation's context.

The information discourse will capture and record several types of stories. One type will represent a record of the course of the analysis and the communication between the user and the system. This type of story will present the basic methods by which we formulated hypotheses about the world, articulated the new knowledge, and communicated with others. Another type of story will represent the product of the analysis and contain a record of the resulting conclusions and new knowledge. We envision presenting this story as a digital media-enhanced communication that tells how the conclusions from the analysis were made.

Achievement of this requires that analysis methods are flexible and the results can be reused for many different user operations. Information about the type of analysis, the requirements and preferences of the user, and the pedigree of the data will be required to provide the user with personalized information spaces that change as the user's needs change.

In order for the system to make deductions that help the user arrive at conclusions from the knowledge presented in the visualizations, additional information must be recorded. A base of knowledge describing the relationships among different aspects of the data must be maintained in a way that allows automated manipulations and inferences to be performed. In fact, the new information discourse must function both as a sophisticated visualization environment and as an intelligent information system.

Our new generation of visual analysis tools will support cooperative analysis by teams of users who share information with one another, both in a shared working environment around a single visualization and through shared knowledge models used by the individuals working alone. One of the keys to accessing and communicating "larger perceptions" in the digital age may well lie in the collective social activities that occur within the context of sharing these stories. We can envision collaborative narratives within this information discourse. For example, imagine an analyst being at a dead end and building a story representing the current state of the analysis to share with a colleague. In discussions with the analyst, the colleague adds insight and knowledge to the story, which puts the original analyst back on track. Exchanges and experiences in group exploration and discovery (communal "curious learning") promise rewards far beyond the mere story.

As we move towards such higher-order interaction methods for data-intensive computing, solutions will require interdisciplinary thinking and problem solving. A variety of exciting research lies ahead, as we develop:

- New signatures (i.e., a mathematical representation of the object that captures its features and their strengths) for diverse data, including numerical data and diverse media.

- Methods for representing contextual information (and associated signature), including models of analyst knowledge and insight, collection characterization and summarization, and state-of-the-world information.
- New visual paradigms that reflect collection and situational context and analyst knowledge.
- Methods for incremental processing of data and fusing information across media and sources.
- New paradigms for interacting with information and with other analysts.
- New paradigms for dealing with scale and complexity.
- New approaches for reasoning with loosely controlled or semi-structured data and models
- New human-computer interaction capabilities that reduce the requirement for keyboard/mouse input from the user
- An improved understanding of how humans arrive at conclusions and make generalizations, subsequently improving our ability to teach machines to do the same.

Information discourse has evolved from data presentation to visualization to data mining to the state of the art - visual data mining. We expect that the enormous growth in size and variety of information will continue, driving requirements for new computing systems, new user interfaces, new applications, and what comes after visual data mining.

5 Conclusion

The new human information discourse we have described extends current visual analysis capabilities to become a visualization-based, knowledge discovery environment that works with – not for – the user. This two-way, interactive dialogue enables the system to respond appropriately, learning from and aiding the human in the discovery process and allowing the user to develop and tell a story that represents the knowledge and learning that has been gained. Much exciting research is required to reach this vision but the payoff will be enormous. This new human information discourse will enable even better quality and quicker knowledge discovery that will help us keep ahead of the growing body of data and information that currently threatens to overwhelm us.

References

[Ahlberg & Shneiderman 1994] Ahlberg, C. and Shneiderman, B.: “Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays”; Proc. CHI ’94 Human Factors in Computing Systems, Ed. Adelson, Beth; Dumais, Susan, and Olson, Judith. ACM Press/Addison Wesley (1994). 313-317, 479-480.

[Card et al. 1991] Card, S., Robertson, G., Mackinlay, J.: “The Information Visualizer: An

Information Workspace"; Proc. CHI '91, ACM Conference on Human Factors in Computing Systems (1991), 181 – 188.

[Card et al. 1999] Card, S., Mackinlay, J., and Shneiderman, B. (Eds.): "Readings in Information Visualization: Using Vision to Think"; Morgan Kaufmann Publishers, Inc/ San Francisco (1999).

[Fairchild et al. 1988] Fairchild, K., Poltrock, S., and Furnas, G.: "SemNet: Three-Dimensional Graphic Representations of Large Knowledge Bases"; in Guindon, Raymonde (ed.) Cognitive Science and its Application for Human-Computer Interaction. Lawrence Erlbaum and Associates/ Hillsdale, New Jersey (1988), 201-233.

[Gershon & Eick 1997] Gershon, N. and Eick, S.: "Information Visualization"; IEEE Computer Graphics and Applications, July/August (1997), 29-31.

[Havre et al. 2000] Havre, S., Hetzler, B., Nowell, L.: "ThemeRiver: Visualizing Theme Changes Over Time"; Proc. IEEE Symposium on Information Visualization, InfoViz 2000 (2000), 115 – 123.

[Hearst 1995] Hearst, M.: "TileBars: Visualization of Term Distribution Information in Full Text Information Access"; Proc. CHI '95 (1995), 59-66.

[Hemmje et al. 1994] Hemmje, M., Kunkel, C., & Willett, A.: "LyberWorld - A visualization user interface supporting full text retrieval"; Proc. SIGIR '94 (1994), ACM, New York, 249-259.

[Jain & Dubes 1988] Jain, A. and Dubes, R. (1988) Algorithms for Clustering Data. Prentice Hall.

[Keim & Kriegel 1996] Keim, D., Kriegel, H.: "Visualization Techniques for Mining Large Databases: A Comparison, Knowledge and Data Engineering"; IEEE Transactions, v. 8 (1998), 923-938.

[Lin 1992] Lin, X.: "Visualization for the Document Space"; Proc. Visualization '92 (1992), 274-281.

[Lyman & Varian 2000] Lyman, P. and Varian, H.: "How Much Information"; retrieved from http://www.sims.berkeley.edu/how-much-info on (May 2001).

[McCormick et al. 1987] McCormick, B., DeFanti, T., Brown, M.: "Visualization in Scientific Computing"; *Computer Graphics*, 21, 6 (Nov. 1987), 1 – 14.

[Miller et al. 1998] Miller, N., Wong, P., Brewster, M., Foote, H.: "TOPIC ISLANDS™ - A Wavelet-Based Text Visualization System"; Proc. IEEE Visualization '98, ACM Press (1998), 189-196.

[Nowell et al. 1996] Nowell, L., France, R., Hix, D., Heath, L., Fox, E.: "Visualizing Search Results: Some Alternatives to Query-Document Similarity"; Proc. SIGIR '96, ACM Press, Zurich (1996), 67-75.

[Olsen et al. 1993] Olsen, K., Korfhage, R., Sochats, K., Spring, M., Williams, J.: "Visualization of a Document Collection: The VIBE System"; Information Processing and Management 29, 1 (1993), 69-81.

[http://www.pnl.gov/infoviz/] PNNL Information Visualization Web Site, http://www.pnl.gov/infoviz/ (2001).

[Rao & Card 1994] Rao, R. & Card, S. K.: "The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information"; Proc. CHI '94, 318-322. New York: ACM Press (1994).

[Rasmussen 1992] Rasmussen, E.: "Clustering Algorithms"; in Frakes, W. & Baeza-Yates, R. (eds.), Information Retrieval: Data Structures and Algorithms, Prentice-Hall, Englewood Cliffs, N.J. (1992), 419-442.

[Rbarsky et al. 1999] Rbarsky, W., Katcz J., Jiang F., Holland, A.: "Discovery Visualization Using Fast Clustering"; IEEE Computer Graphics and Applications, v. 19 (1999), 32-39.

[Rich & Knight 1991] Rich, E., Knight, K.: "Connectionist Models"; in DM Shapiro and JF Murphy (eds), Artificial Intelligence, McGraw Hill, New York (1991), 487-528.

[Russell & Norwig 1995] Russell, S., Norwig, P.: Artificial Intelligence, Prentice Hall, Englewood Cliffs, NJ (1995).

[Schank 1995] Schank, R.: "Tell Me a Story: Narrative and Intelligence"; Foreword by Gary Saul Morson, Northwestern University Press, Evanston, IL (1995). Series title: Rethinking theory.

[Spoerri 1993]. Spoerri, A.: "InfoCrystal: A Visual Tool For Information Retrieval"; Proc. of Visualization '93, ACM Press, Los Alamitos, CA (1993), 150-157.

[Thomas et al.] Thomas, J., Cook, K., Crow, V., Hetzler, B., May, R., McQuerry, D., McVeety, R., Miller, N., Nakamura, G., Nowell, L., Whitney, P., Wong, P.: "Human Computer Interaction with Global Information Spaces - Beyond Data Mining"; Proc. British Computer Society Conference (1999).

[Tufte 1983] Tufte, E.: "The Visual Display of Quantitative Information"; Graphics Press, Cheshire, CT (1983).

[Tufte 1997] Tufte, E.: "Visual Explanations: Images and Quantities, Evidence and Narrative"; Graphics Press, Cheshire, CT (1997), 90-91.

[Tufte 1990] Tufte, E.: "Envisioning Information"; Graphics Press, Cheshire, CT (1990).

[Ware 2000] Ware, C.: "Information Visualization: Perception for Design"; Academic Press, San Diego (2000).

[Wong 1999] Wong, P.: "Visual Data Mining"; IEEE Computer Graphics and Applications, IEEE CS Press, Los Alamitos, CA, 19, 5 (1999), 20 – 21.