VizCept: Supporting Synchronous Collaboration for Constructing Visualizations in Intelligence Analysis

Haeyong Chung*, Seungwon Yang, Naveed Massjouni, Christopher Andrews, Rahul Kanna, and Chris North

Department of Computer Science, Virginia Tech

ABSTRACT

In this paper, we present a new web-based visual analytics system, VizCept, which is designed to support fluid, collaborative analysis of large textual intelligence datasets. The main approach of the design is to combine individual workspace and shared visualization in an integrated environment. Collaborating analysts will be able to identify concepts and relationships from the dataset based on keyword searches in their own workspace and collaborate visually with other analysts using visualization tools such as a concept map view and a timeline view. The system allows analysts to parallelize the work by dividing initial sets of concepts, investigating them on their own workspace, and then integrating individual findings automatically on shared visualizations with support for interaction and personal graph layout in real time, in order to develop a unified plot. We highlight several design considerations that promote communication and analytic performance in small team synchronous collaboration. We report the result of a pair of case study applications including collaboration and communication methods, analysis strategies, and user behaviors under a competition setting in the same location at the same time. The results of these demonstrate the tool's effectiveness for synchronous collaborative construction and use of visualizations in intelligence data analysis.

KEYWORDS:. Collaborative visualization, text and document data, intelligence analysis.

INDEX TERMS: H.5.3 [Group and Organization Interfaces]: Collaborative computing, Computer-supported cooperative work, Web-based interaction

1 INTRODUCTION

The analysis of large, complex abstract data is becoming essential in various problems of national security, disaster and crisis management, customs control, and so on. These analytic and decision-making processes over vast amounts of information involve sensemaking, a process in which information is collected, organized, and analyzed to generate a productive conclusion and to initiate new lines of inquiry or questions [1].

A wide variety of visual analytics tools have been developed to support the sensemaking process. Primarily developed for single users, these tools provide support for searching, filtering, isolating, visualizing, and identifying connections in a large aggregate of data. However, the size of some intelligence data sets is very large and the volume often exceeds a single user's analysis and exploration capability. In addition, a single person analysing a

*e-mail: chungh@vt.edu

large amount of data may be inefficient, and may potentially introduce individual bias into the results. This work is motivated by the prospective benefits of synchronous collaborative creation of visualizations and the combination of individual workspace and shared visualization in the analysis of a large quantity of intelligence data. Our design focuses on how both individual workspaces and shared visualizations can be structured in order to lead to an effective collaborative environment and what kind of design space and approaches might support this collaboration environment efficiently.

From these design considerations, *VizCept*, a web-based collaborative visual analytics tool, was developed. VizCept provides both individual analysis space and global visualization of group findings in a collaborative environment. The synchronous collaboration supported by VizCept saves analysis time by efficiently distributing multiple analysts' tasks in real-time, since individual analytic results from each user's workspace are simultaneously aggregated in a global concept map [2].

To support synchronous collaboration in analytics tasks, VizCept generates global visualizations immediately when users add entities and relationships, and allows the users to refine the visualizations through collaboration. A distinguishing feature of VizCept is the support of multiple techniques to maintain awareness among collaborating users. User awareness of shared information is critical for multiple users to have fluid collaboration [3]. However, this aspect of collaboration has been infrequently explored in the visual analytics domain. Awareness enables users to understand other participants' activities and presence without the interruption of the main task and enables the users to solve collaborative problems fluidly [4]. In addition, awareness enables the user to coordinate their work more efficiently during their analytic efforts. With awareness support, collaborating users are able to grasp what other users are working on and where their focuses are directed. This, in turn, can promote more focused and productive communication between analysts.

To demonstrate the effectiveness of synchronous collaboration in creating visualizations in VizCept, we have applied our tool to VAST contest datasets under different environments. We found that the tool was able to help analysts to find important information, such as key persons, organizations, and relationships, through clusters of information on the concept map that formed as the result of the synchronous collaboration.

The following sections discuss related work and describe the three main views, the design considerations made for synchronous collaboration, and the implementation of the internal logic of the tool. Then we present a use case of VizCept. We discuss user behavior and interesting collaborative analysis strategies using the tool. Finally, we discuss limitations and potential future features.

2 RELATED WORK

VizCept is inspired by previous work in three areas - collaborative concept mapping, visual analytics, and Web-based collaborative visualization tools.

Concept mapping has been broadly studied and used in knowledge representation and teaching/learning areas due to its effectiveness in externalizing thoughts as abstract concepts and relationships. Computer and Web technologies have led to the development of concept mapping in collaborative environments using networked computers. For example, CMapTools supports collaboration by allowing users in remote locations to develop a shared concept map [5]. Representation 2.0 supports the collaborative construction of semantic representations through the active engagement of its users in the educational context [6].

In the visual analytics domain, four tools are closely related to VizCept. These tools focus on representing the relationship among entities extracted from documents or other textual data in a semantic graph or different types of visualization. Jigsaw provides multiple views that visually illustrate connections between automatically generated entities in multiple documents [8]. In this tool, the views are interoperable, and changes in a view are seamlessly reflected in the other views. The multiple views allow a user to explore the various connections between entities and documents within the data set. Analyst's Notebook by i2 Inc. provides semantic graph visualization for link analysis to identify connections and patterns in a large aggregate dataset [9]. It allows users to visualize and analyze large quantities of intelligence data through basic link and node charts.

The third tool, Entity Workspace, also follows this basic paradigm. It was modelled on the effectiveness of a traditional evidence file that keeps track of various facts about entities and relationships, such as people, places, organizations, telephone numbers, bank accounts, etc. [7]. It provides an explicit model of important entities by allowing users to find potentially important documents and entities. This tool helps analysts find the new facts based on these connections of entities and information rapidly. In their follow-up work on Entity Workspace, Bier et al. developed five design guidelines for collaboration for intelligence analysis and modified the Entity Workspace system based on these guidelines [10]. Like VizCept, the tool helps collaborators to connect entities and concepts that are found by different analysts, allowing them to be merged and shared seamlessly.

The fourth tool also directly addresses collaboration. Pike et al developed a service oriented visual analytics system, SRS [11], which decouples analysis and visualization of a dataset to allow client applications running on different devices, such as mobile devices, to access the analytic service and effectively render the provided data. The SRS web client incorporates web services similar to those provided by VizCept, such as manually created concept maps, timeline visualizations, and listings of query results. It also allows users to save and share questions, hypotheses, evidence, etc.

Both Entity Workspace and SRS have some similar features to VizCept that make it possible for multiple analysts to work on the same analytical task, but the communication and collaboration techniques used to connect collaborators are different from VizCept. These tools are primarily designed for asynchronous collaboration, which directly influences how these tools support awareness and workflow.

Another area related to VizCept is the development of systems for Web-based collaborative visualization. Many of these collaboration tools focus on group exploration of data through community components such as annotations and comments from remote environments. They emphasize the engagement of users in data analysis through various types of social navigation like associated discussion components where users post comments or annotations, and ask questions.

In many cases, one user initially creates the visualization, and other users add or annotate visual data or show interesting views for data. For example, *ManyEyes* allows users to upload their data onto a public website and build, share, and edit information visualizations from the data [12]. Another example is Dashiki, a wiki-based website which enables multiple users to build wiki-based visualization dashboards through a user-editable wiki mark-up language and interactive editors [13]. Users can present and organize their own dashboards, which contain visualization and presentations that are created by multiple users as community components. Heer et al. developed a web-based asynchronous collaboration visualization tool, *sense.us* [14]. It provides a set of interactive visualization features along with collaboration via bookmarking of views, a new discussion scheme called "doubly-linked", graphical annotation, and social interaction through annotations, comment listings and user profiles. Increasingly, many web-based services are supporting more collaborative visualization features for general web users. For example, Google Maps supports a personalized map service that allows users to share specific location information with others [15].

3 VizCept System Overview

In this section, we discuss the features and design approaches of our tool, focusing on its user interface. VizCept consists of three main views as shown in Figure 1. They are the *workspace*, the *concept map* view and the *timeline* visualization. In addition, raw data, such as text, images, and spreadsheets can be opened in new windows through links provided directly in the workspace. Individual users are able to work on any component to analyze the data that they are interested in, and the analyzed data is aggregated automatically in the Concept map and Timeline views.

The workspace (Figure 1a) is the main component for dataset exploration, providing keyword searching and file content browsing. All of the queries issued by the user are shown in the keyword list located at the top right corner, and below that is the relationship tool, which allows the analyst to specify relationships between keywords. As keywords and relationships are added, they are automatically added to the concept map view (Figure 1b) and the timeline visualization (Figure 1c).

The workspace is an environment specific to each analyst, while the concept map and timeline visualization are spaces formed by the aggregation of all keywords and relationships from all participating analysts. This combination of individual and shared spaces allows for the distribution of dataset exploration tasks among analysts and integration of findings into a unified plot thereby promoting improved efficiency.

3.1 Workspace

The Workspace (Figure 2) is the main entry point into the system and the information foraging tool. In this space, analysts find specific information by reading or skimming each relevant data file based on a keyword search. The view can be considered a detailed, local view of the data, providing an individual workspace to search for keywords in the dataset, create concepts and assign relationships between the concepts.

Each user has an individual workspace, but read-only access to the other users' workspaces is available through tabs located above the search box (Figure 2d). This allows users to view the searches and relationships that others are making and gain some insight as to what direction others are pursuing by comparing concepts and relationships through detailed textual information.

The search box (Figure 2b) accepts regular expressions for detailed custom search. Keyword search is the basic but essential method for retrieving specific information from a large textual
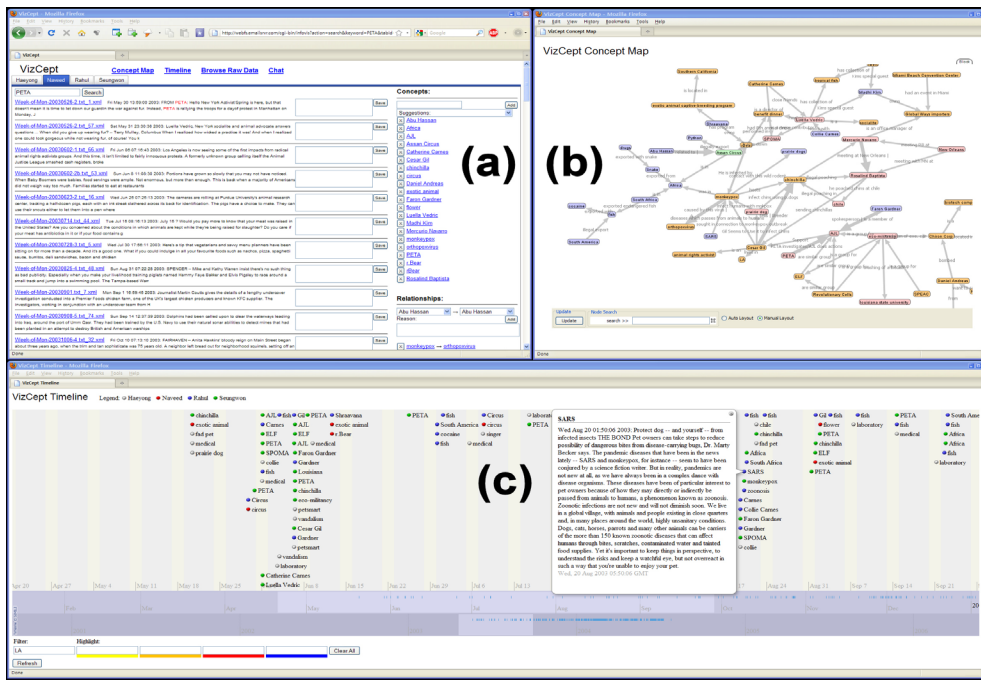
Figure 1. Three main views of VizCept: (a) WorkSpace, (b) Concept map view and (c) Timeline view.

dataset. Each row on the search results page shows a link to the document that includes the keyword, content of the document, and a note box to enter shared comments (Figure 2e).

Users can hover their mouse over the document links to preview the content for quick browsing as shown in Figure 2a. Note that the current search term is highlighted in red (Figure 2c) for easy identification. If a document looks interesting, the user can click on the link to open the document in a new window.

Notes are shared, so once a note has been added to a document, it will be visible to all users when they encounter the document. This allows analysts to communicate important thoughts about specific documents and to provide clues about documents that have been previously viewed.

Users add concepts to the workspace in three different ways. First, each keyword a user searches for is automatically added as a concept in the concept list at the top right corner (Figure 2f). Second, users can manually type in the concept box. Third, users are provided with concept suggestions based on the existing concepts added by other users in their own workspace (Figure 2i). This list helps analysts avoid creating redundant concepts. While the system automatically matches concepts based on text, it does not handle issues such as alternate spellings or stemming (e.g., 'butter fly' and 'butterfly'). In this case, concepts will not be associated and connections may be missed. To avoid this problem, users can select suggested keywords and add them to their workspace from the concept suggestion list directly.

To develop a plot, connecting concepts with meaningful relationships is essential. The relationships specified by the analyst are listed in the relationship list (Figure 2g). Relationships are specified through two popup menus containing all of the concepts identified by the analyst. Once the concepts are selected, the analyst can enter a phrase that describes the relationship between the two concepts into the 'Reason' text box. When the 'Add' button is clicked, the two concepts and the reason for the relationship are displayed on the bottom right corner of the screen. The relationship and concept can be removed with a delete button marked with 'X'.

From the workspace, users also can launch the concept map view, the timeline visualization, chat windows, and access raw data files by following the corresponding links at the top of the workspace (Figure 2h). The 'Browse Raw Data' link at the top of the screen directs the user to a list of data files for further document analysis.
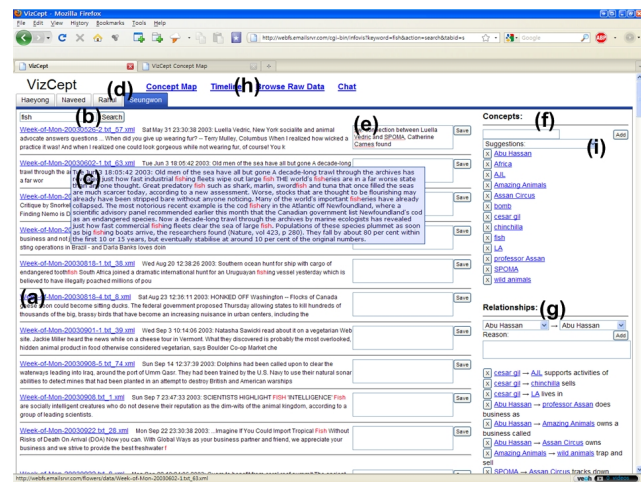

Figure 2. Workspace user interface: (a) document links, (b) search box, (c) preview the document (d) user tabs, (e) note box, (f) box to add a concept (g) box to create a relationship, (i) suggestion of concept, and (h) links to Concept map, Timeline visualization, raw dataset access and chat tool.

3.2   Concept Map View

The Concept Map (Figure 3) displays the union of all the concepts and relationships that each user has discovered in their own workspace. In this view, users can easily keep track of the concepts added by other users and progressively make connections between them to make sense of the relationships. The

nodes in the visualization represent entities such as names, locations, objects, and concepts, while relationships are represented as directed edges, labeled with descriptive titles. The node colors represent the color of the user who first specified the concept. As there can be multiple relationships between a pair of entities, the edge thickness increases as multiple relationships are added to the same concept pair by different analysts.

In support of collaboration, sharing this visualization among analysts can play an important role in generating new insights and hypotheses. It can help track valuable information, and aid in reaching a conclusion. The global concept map allows analysts to understand the entire plot and may help them gain greater insight. Various interaction methods, such as pan/zoom or manual/automatic reorganization of the layout of the map, are supported by the concept map view to help the analyst explore the visualization.
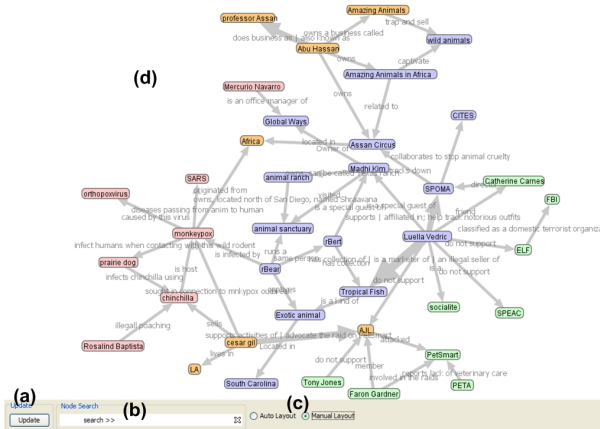
Figure 3. Concept map web applet: (a) update button, (b) search for keyword, and (c) layout selection (manual or automatic), (d) display area.

The concept map view is built in a web applet. It consists of the concept map display area (Figure 3d), an update button (Figure 3a), a search panel (Figure 3b), and layout selection options (Figure 3c). The search panel allows the user to highlight a specific node in the concept map view by typing a word into the search box. The "Update" button updates the nodes and edges dynamically by reading the newest data from the server. VizCept's primary layout technique is the force-directed graph algorithm provided by the Prefuse toolkit [16]. Optionally, the user can turn off the automatic layout and manually adjust the arrangement of the graph.

An important consideration in collaborative interfaces is conflict resolution or reconciliation. A potential issue in VizCept is the addition of multiple labels to the same relationship link. VizCept supports the addition of multiple labels by displaying both terms separated by '|' (Figure 4). Therefore, users can identify relationships that have conflicts and are able to resolve conflicts. This allows the users to resolve conflicts through social and explicit communication methods, such as observation of other people's workspace, negotiation, discussions, and Internet chats. This conflict resolution approach is preferred in collaborative visualization systems [17, 18].

The concept map view supports several exploration awareness features for collaborating users' activities and presence, such as color-coding, edge thickness and multiple edge labels. All of these techniques are visible in Figure 4. In this example, we can see that three users have been contributing to the concept map. All three users have added relationships and labels between the 'Luella

Vedric' and 'Tropical fish' concepts ('has collection of', 'is a marketer of', and 'an illegal seller of'). This makes the edge three times thicker than that of single relationship nodes. This thickness becomes a potential indicator of the significance of the relation, since many users have found it to be important.
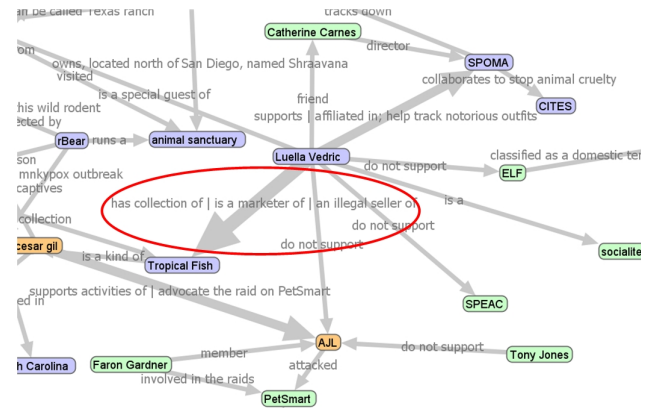
Figure 4. Variable edge thickness and multiple edge labels between two nodes: thickness of edge can gives some different weights for a specific relationship and these relationships show stronger interest and importance. Nodes are color-coded for each user.

3.3 Timeline View

Identifying and presenting date/time-related information plays an important role in developing a story from the dataset. That information can assist in describing the context of an entity. Since the concept map view focuses more on the relationships among important entities and lacks the date/time information, the timeline view has been developed to compensate this limitation.

For the timeline, we used the SIMILE timeline widget [20] (Figure 1c). Concepts appear on the timeline as events, with the document that connects it to the associated point in time available in the detailed view (Figure 5). Concepts that are associated with multiple documents will appear in multiple locations in the visualization. To support collaboration in this view, the concepts are color-coded based on the user who added the concept.
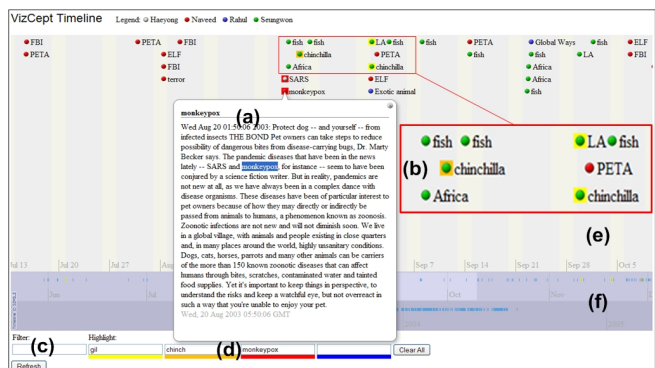
Figure 5. Timeline visualization with upper band (for details by months) and lower band (for overview by years): (a) event details, (b) magnified image of concepts by users (green or red icon) and concepts with highlight (yellow and orange) (c) 'filter' interaction, and (d) 'highlight' interaction. (e) Upper band (f) Lower band.

3.4 Additional Features

To support collaboration for remotely located analysis sessions, a simple chat interface is provided. The chat tool allows users to broadcast messages to all other users, or selectively send a message to a particular user of the system.

4 IMPLEMENTATION

In VizCept, multiple simultaneous messages and requests occur and they should be processed seamlessly without much delay. Communication among the three main components in VizCept is achieved via a web server and a database. Concepts, relationships and notes are all sent back to the server, where they are combined to create the unified data source available to all of the users in the various views.

The Workspace component consists of a web front-end that queries a web server backed by a database containing intelligence data, entity and relationship data in the concept map, and date/time-related data. The front-end provides a mechanism by which users can query and update the data and display the results. The data is exchanged between the client and the server using files in the Java Script Object Notation (JSON) format.

To support an interactive concept map on the web-browser, our concept map view is implemented as a Java Applet using the Prefuse toolkit [16]. The concept map view supports various interaction methods including panning, zooming, and node drag-and-drop. The underlying concept map data is communicated using GraphML [19] and XML retrieved from the VizCept's backend server in real-time. GraphML is used so that VizCept can support the special graph features such as the colored nodes and varying edge thicknesses.

The timeline visualization is implemented with the SIMILE [20] timeline widget, which utilizes a Javascript API. When the timeline view is refreshed, it contacts the server, which generates an XML file based on the current state of the system. Filtering and highlighting of events is done on the client side by reading the Document Object Model (DOM) object and updating it by finding the events in the timeline and removing or highlighting them based on the user action.

5 USE CASES

We carried out two systemic case study applications of VizCept. The goals of the two applications were to see if VizCept can accommodate analytic tasks for different sized datasets and synchronous collaborative environments, to learn how users can collaborate using this tool, and to demonstrate the effectiveness of collaborative creation of visualization along with user awareness. In both of the case study applications, the four developers of the tool collaborated to solve analytic problems using VizCept.

5.1 VAST Datasets and Application Environments

The case study applications involved the use of two different VAST contest datasets: Blue Iguanodon [21] and a modified version of White Smilodon [21]. Both datasets present scenarios related to eco-terrorism and suspicious illegal activities involving endangered animals.

The first session was essentially a longitudinal analysis intended to evaluate the tool and the supported collaborative process. The analysis was carried out by the four team members in a time-unlimited laboratory environment using the Blue Iguanodon dataset. All four team members were graduate students in Computer Science enrolled in a graduate level Information Visualization class. This dataset consists of approximately 1700 documents including more than 1500 news stories, various textual documents, pictures, and spreadsheets.

The second case study application of VizCept was to participate in a live competition. This contest followed the format and structure previously described by Whiting et al. [22]. During this competition, 38 graduate and undergraduate students competing in teams of four to five all raced to analyze a modified version of the White Smilodon dataset in a two-hour period. Each team competed with its own custom-built visual analytics tool. There were five graduate student teams from the Information Visualization class, and five undergraduate teams from an Introduction to Human Computer Interaction course. Before the contest, all of the graduate students teams had previously solved the Blue Iguanodon scenario using their respective tools, and the undergraduate students teams had solved the Stegosaurus scenario [23]. The White Smilodon dataset, which was modified by the organizer of the contest to fit the two-hour time limit, contained approximately 500 documents, including several slides and pictures from the original set.

5.2 Longitudinal Analysis

In this case study application, the team members developed an analytic workflow consisting of three phases: initial work-division, foraging, and sensemaking (Figure 6). The following sections discuss each of these stages and how they were supported by VizCept. One team member used a desktop PC while the others used laptop computers. The four team members collaborated to identify the story hidden in the dataset without a time limit by creating visualizations with VizCept. All of the team members conducted the analysis in the same location and at the same time.

During this case study application, all screen activity was recorded using Camtasia [24], and all concepts, relationships and notes were recorded into a database and retained. Observations were written down as they were made during the session. A debriefing followed the session and further observations about the experiences of performing the analysis were discussed among the team members and written down after the session. Further discussion with the other authors led to the following description of the analysis process.

5.2.1 Initial Work-Division Phase

The Blue Iguanodon dataset was preloaded in VizCept before the analysis work began. The analytic process started with each team member working on his or her own analysis in individual workspaces.

The seed document of the Blue Iguanodon dataset was provided as a starting point for analysis. The document includes backgrounds, tasks, and some advice related to types of data files and analysis methods. All of the collaborating users read the seed document, discussed it, and identified the different seed concepts (Figure 6a, b). Each user then took a couple of seed concepts and performed a preliminary analysis to generate additional related concepts (Figure 6c, d). Steps (c) and (d) were repeated until 5-10 concepts were found. To avoid redundancy of analysis, the users then coordinated by dividing and then assigning disjointed clusters of potentially meaningful and suspicious keywords to each user (d). This initial assignment of clusters was conducted through face-to-face negotiation based on the user interests.

5.2.2 Information Foraging Phase

Individual document analysis. Based on these initial keyword sets, the team members started performing searches and reading files using the VizCept workspace (Figure 6d). In this step, new concepts and relationships were identified by reading each
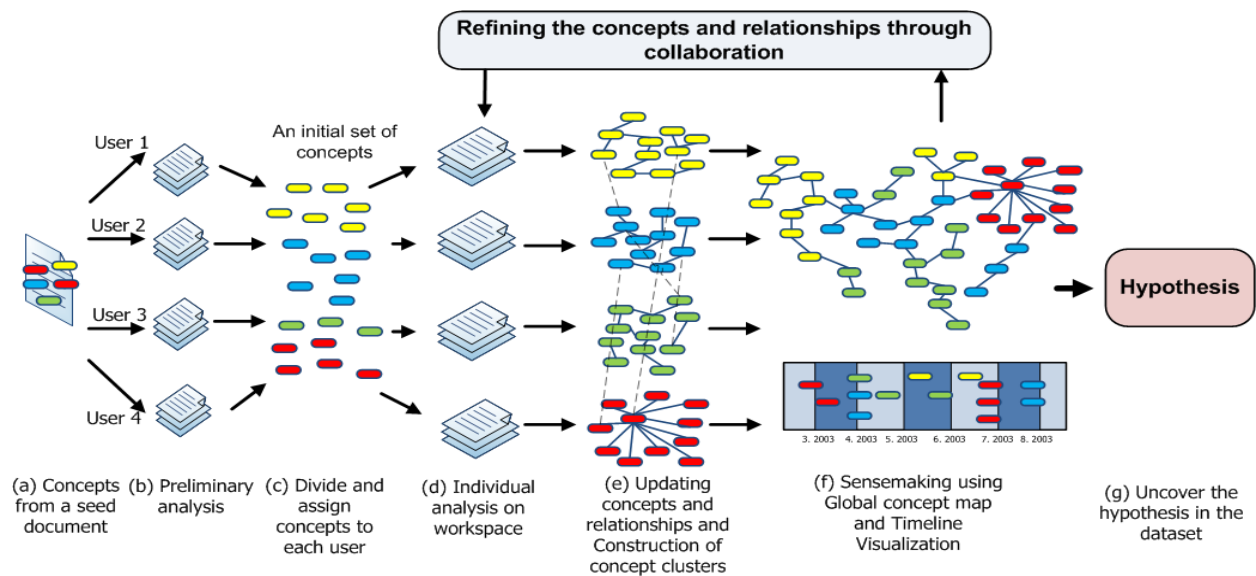
Figure 6. VizCept analysis workflow for synchronous collaboration.

document. In addition, the users consulted the other types of available data, such as images or spreadsheet files in the dataset to identify missing connections. When users found important information related to a specific document, they left notes linked to the document for the other users.

Updating concepts and constructing concept map clusters. Building out from the initial cluster of concepts assigned to them, each user continued exploring the dataset, identifying more concepts and making more relationship labels between the concepts (Figure 6e). The disparate lines of investigation meant that initially, the work of each user existed as a disconnected sub-graph in the concept map.

However, as the investigation progressed, these concept clusters were expanded and connections began to form between the sub-graphs (Figure 6e, f). At this stage, the importance of the concept map increased as it began to show links between the separate investigations. Nodes with high connectivity highlighted important concepts and thicker links showed the increased interest and significance of a relationship between concepts. Conflicting findings and potentially incorrect information were discussed through short face-to-face meeting or online chat. The concept map also helped to identify less important or redundant concepts, which were discussed and occasionally removed from the system.

As users added concepts, they first checked the suggestion list in the workspace to see if the concept had been created already by other users to reduce redundancies. Overlapping or preexisting terms indicated that one of the other users had already shown an interest in the concept. Discovery of these frequently led to the user checking the concept map to see which user added and then initiating a conversation to resolve conflicts or discuss the investigative paths leading to the concept. If a document file was found to be valuable to other team members as well, then the users first tried explicit communication with the other team members, who may be interested in it. Often this was followed by sending the file link to the other team members via the chat view.

An interesting variant of this process was demonstrated by one of the users who leveraged the collaborative nature of the tool to provide serendipitous connections. Rather than carefully choosing concepts that seemed particularly interesting or important, these users added all of the keywords that seemed remotely related to the documents they were reading. This made quick connections to

the work of other users, thus highlighting the important leads to follow. Of course, while this behavior directed his investigation, it came at the cost of polluting the concept map with marginally relevant noise.

5.2.3 Sensemaking Phase

Analysis using visualizations. The primary visualization was the concept map. The users tended to use a mix of automatic and manual graph layout in the concept map – manually arranging the sub-graph they were interested in, using the rest of the graph to provide contextual information about the progress of the other users and to see connection between their work and the work of the rest of the group (Figure 7). Due to the nature of the dataset, the timeline was less useful, but it was used for a couple of specific tasks, such as examining concepts like 'monkeypox' to see when the outbreak was first noticed (i.e., simply pick the left most occurrence of the dot-and-concept pair) and to identify the first people to become infected (i.e., browse through dot-and-concept pairs to find names). As with the concept map, users used the different colors that are assigned to different users and the filtering mechanism to filter events in the timeline. By using the filtering technique in the timeline, the users were able to narrow down to particular concepts and events.

Refining concepts through collaboration. The primary tool for refining concepts was again the concept map visualization. By keeping track of the nodes and links on the concept map, the users could detect key players (i.e., many links were connected to this node) and how they were connected to other organization or events. The users could also distinguish potentially unrelated or unimportant information, which was rarely linked to other clusters of the global concept map

When users found interesting information in the concept map, they frequently spoke directly to the other user who created that concept by checking the corresponding user's assigned color. Explicit communication was important in negotiating the different thoughts on the plot. For example, some team members argued that Luella Vedric was an animal rights activist based on multiple labels on relationships, but another member claimed that she was not only an animal activist, but also a secret marketer of illegal animals. He added another label "marketer of illegal animals" and provided related concepts and relationships to the other team

members. After a short discussion, the team reached a consensus that Luella was a socialite and an animal rights activist, but also an illegal marketer of endangered animals. Following these patterns, the users iteratively refined the concept map by adding or removing concepts and relationships through face-to-face discussions and finding new information.
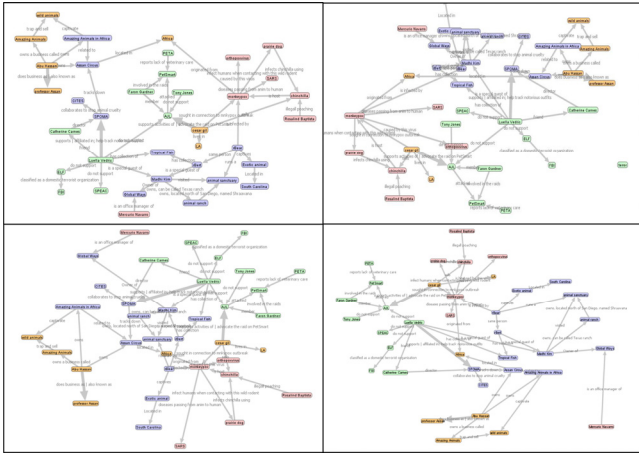

Figure 7. User generated concept map layouts for each user during the analysis.

5.2.4   Converging on a Story

The concept map view and timeline visualization helped the users to converge on a common understanding of the dataset. When growth of the concept map was showing signs of halting, and all of the team members agreed there was nothing more to add, the users met to develop a global layout of the concept map. By taking turns, each user explained his cluster in the concept map, adjusting the layout of the cluster in turn. Brief discussions and conflict resolutions happened during this time. When everyone finished his or her explanation, a global layout was prepared along with a story that described the hidden plot in the dataset.

By the end of the five-hour analysis, the group had successfully uncovered the hidden plot in the Blue Iguanodon dataset as determined by an external scorer. The final shared data space contained 118 concepts, of which 35 were disconnected singletons and the remaining 83 were connected with 98 relationships.

5.3   Live Competition

The primary goal of participating in the live contest was to illuminate usability issues and process support that would be exacerbated or highlighted by the time pressures of the event. The format of the competition did not permit in situ recording or observation, but the database records were retained and a post-session debriefing was held to discuss the experience. The notes and observations from this session were then discussed with the other authors, leading to the following key observations.

Since the same team was assembled for the live competition, the analytic approach followed the same general pattern as the one developed in the previous case study application. The primary change was an increase in communication and a greater level of collaboration. Anecdotally, all of the users report that verbal communication increased in the live environment. Another potential sign of the increase in collaboration was the proportional increase in the number of notes added to the document collection as annotations. In the longitudinal analysis, 53 of the 1700 documents were annotated during the five-hour session. During the live competition, 31 of the 400 documents were annotated in

the two-hour session. If we normalize this in terms of the duration of the sessions, this is 10.6 notes an hour for the first session and 15.5 notes per hour in the second, or a 46% increase. While this may be due to the dataset, it is suggestive that more sharing of information was occurring.

The team members also found that they monitored the progress of their partners through the concept map more closely during the live session. Rather than concentrating purely on individual investigations, more emphasis was placed on lines of investigation that tied back into the work of the group. This increased awareness helped the group as a whole maintain focus. In one notable instance, some of the users noticed a growing sub-graph in the concept map that was not tied back to the rest of the investigation. They discussed this with the user producing it and steered him back towards a more productive line of investigation.

By the end of the investigation, 65 concepts had been identified, of which 21 were singletons and the remaining 44 were connected by 58 relationships. Of the ten teams that participated in the contest, the VizCept team earned the highest overall solution score based on the metric used in [21] and was one of two teams that correctly identified all sub-plots. At a post-contest session, Vizcept was reviewed by the other contest teams, where it won the popular vote for best analytic tool. In the reviews, they mentioned that VizCept's approach to the synchronous collaborative process best supported the small-team task scenario. In contrast, most of the other teams' tools focused specifically on visualization or interfaces for foraging rather than on process, making collaborative, time-pressured analysis difficult.

6   DISCUSSION

The synchronous collaboration supported by VizCept appeared to be a successful approach to collaborative visual analytics. The shared concept map and various communication channels allowed the users to pursue individual lines of investigation, while still maintaining awareness of the activities of the other users, eventually connecting and establishing a shared understanding of the data. However, there are issues that could still be explored.

One of the challenges in designing a system that supports synchronous collaboration is the management of the tradeoff between collaborative activities and individual activities. There are a number of considerations to be balanced. For example, we wish to foster a high level of awareness, but not at the expense of interfering with individual work. VizCept manages this by making updates to the views of the shared space be manual. This allows the user to work in the visualization without having it constantly changing, but, as was discovered in the case study applications, this can come at the cost of awareness. Another issue, not currently addressed by VizCept is the issue of privacy. Analysts may not wish to share their results prematurely, or they may wish to share selective portions of their investigation. VizCept shares everything instantly and indiscriminately. The incident during the live competition showed one of the benefits of this, but redirecting the seemingly disconnected analysis could have closed off a promising line of investigation as well or created social discomfort.

The case study applications also demonstrated that the interface of the workspace could be improved. There are too many steps to create relationships and the manual mechanism for checking for conflicting concepts needs to be more transparent. For example, auto-completion of search terms would be a helpful indicator of preexisting terms.

VizCept provides a number of communication channels, including shared awareness in the visualizations, shared annotations, and chat. In addition, in the case study applications,

the co-presence of all of the users led to a great deal of verbal communication. The chat tool was primarily for passing links that could not be easily conveyed through spoken word. While this arrangement was very effective for collocated collaboration, other avenues of communication would need to be explored for remote collaboration.

Another concern is the scalability of our approach. On a 1280×1024 resolution LCD monitor, the concept map view had a visual limit of approximately 75 nodes on the display to read all concepts, relationships, and labels clearly. However, with 118 entities in the Blue Iguanodon dataset, the important portion of the graph was actually smaller, within the visual scalability range, with many other less-relevant entities and sub-graphs in the periphery. This suggests that such scenarios may naturally break down into manageable chunks. In the timeline view, entity filtering and selective highlighting features helped users to alleviate scalability problems.

7 CONCLUSION

Efficient and accurate analysis of large intelligence datasets requires proper utilization of analytic team members. VizCept is an attempt to improve small-team synchronous analytic collaboration by providing individual workspaces for each collaborating analysts to process the dataset, and shared visualization spaces for teams to generate and share overall insights.

From the small team synchronous collaboration perspective, VizCept helps uncover the hidden story by combining scattered subplots that are found by different analysts. VizCept supports various collaborative features such as color coding of concepts found by each user, concept suggestions found by other users, varying edge thicknesses and multiple labels on edges in the global concept map to express interest by multiple users on that relationship.

Use case results suggest that VizCept may potentially support efficient scalability in large dataset analysis by exploiting small team collaboration. The total amount of analytic work can be divided among the collaborating analysts, thereby reducing the total time to complete the task. Therefore, the workload of each analyst could be reduced. Analyzed information can then be automatically integrated and visualized in a global space for collaborative hypotheses generation. In the case study applications, teams felt that VizCept exploited team collaborative effort better than other approaches, enabling a more thorough exploration of the data and hypotheses.

REFERENCES

[1] S. K. Card, J. D. Mackinlay, and B. Shneiderman, Readings in Information Visualization: Using Vision To Think. San Francisco: Morgan-Kaufmann, 1999.

[2] J. D. Novak, "Concept maps and Vee diagrams: Two metacognitive tools to facilitate meaningful learning," Instructional science, vol. 19, pp. 1-25, 1990.

[3] P. Dourish and V. Bellotti, "Awareness and coordination in shared workspaces," presented at the Proceedings of the 1992 ACM conference on Computer-supported cooperative work, Toronto, Ontario, Canada, 1992.

[4] S. Papadopoulou and M. C. Norrie, "How a structured document model can support awareness in collaborative authoring," presented at the Proceedings of the 2007 International Conference on Collaborative Computing: Networking, Applications and Worksharing, 2007.

[5] A. J. Cañas, et al., "CmapTools: A Knowledge Modeling and Sharing Environment," in The first Int. Conference on Concept Mapping, Pamplona, Spain, pp. 125-133, 2004.

[6] V. Komis, N. Avouris, and C. Fidas, "Computer-Supported Collaborative Concept Mapping: Study of Synchronous Peer Interaction," Education and Information Technologies, vol. 7, pp. 169-188, 2002.

[7] E. A. Bier, E. W. Ishak, and E. H. Chi, "Entity workspace: an evidence file that aids memory, inference, and reading," presented at the IEEE International Conference on Intelligence and Security Informatics (ISI 2006), San Diego, 2006.

[8] J. Stasko, C. Gorg, and Z. Liu, "Jigsaw: supporting investigative analysis trhough interactive visualization," Information visualization, vol. 7, pp. 118-132, 2008.

[9] i2 Inc, Analyst's Notebook. Downloaded 1/24/10. Available: http://www.i2inc.com.

[10] E. A. Bier, S. K. Card, and J. W. Bodnar, "Entity-based collaboration tools for intelligence analysis," presented at the IEEE Symposium on Visual Analytics Science and Technology (VAST '08), Columbus, OH, 2008.

[11] W. A. Pike, et al., "The Scalable Reasoning System: Lightweight visualization for distributed analytics," presented at the IEEE Symposium on Visual Analytics Science and Technology (IEEE VAST), Columbus, Ohio, 2008.

[12] F. B. Viegas, M. Wattenberg, F. v. Ham, et al., "ManyEyes: a Site for Visualization at Internet Scale," IEEE Transactions on Visualization and Computer Graphics, vol. 13, pp. 1121-1128, 2007.

[13] M. McKeon, "Harnessing the Information Ecosystem with Wiki-based Visualization Dashboards," IEEE Transactions on Visualization and Computer Graphics, vol. 15, pp. 1081-1088, 2009.

[14] J. Heer, M. Wattenberg, and F. B. Viegas, "Voyagers and voyeurs: supporting asynchronous collaborative information visualization," presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, San Jose, California, USA, 2007.

[15] Google, "Personalize Google Maps." Avalable: http://maps.google.com/help/maps/mymaps/create.html. Retrieved 3/9/10

[16] J. Heer, S. K. Card, and J. A. Landay, "Prefuse: a toolkit for interactive information visualization," presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, Portland, Oregon, USA, 2005.

[17] P. Isenberg, S. Carpendale, A. Bezerianos, N. Henry, and J-D. Fekete, "CoCoNutTrix: collaborative retrofitting for information visualization," IEEE Comput. Graph. Appl., vol. 29, pp. 44-57, 2009.

[18] S. Huot, C. Dumas, P. Dragicevic, J-D. Fekete, and G. Hegron, "The MaggLite post-WIMP toolkit: draw it, connect it and run it," presented at the Proceedings of the 17th annual ACM symposium on User interface software and technology, Santa Fe, NM, USA, 2004.

[19] GraphML. Available: http://graphml.graphdrawing.org. Retrieved 3/9/10.

[20] SIMILE Timeline Widget. Downloaded 2/21/10. Available: http://www.simile-widgets.org/timeline/.

[21] C. Plaisant, G. Grinstein, J. Scholtz, and M. Whiting, "Evaluating Visual Analytics at the 2007 VAST Symposium Contest," IEEE Comput. Graph. Appl., vol. 28, pp. 12-21, 2008.

[22] M. A. Whiting, et al., "VAST Contest Dataset Use in Education," presented at the IEEE Symposium on Visual Analytics Science and Technology, Atlantic City, NJ, USA, 2009.

[23] P. Saraiya, C. North, V. Lam, and K. Duca, "An Insight-based Longitudinal Study of Visual Analytics," IEEE Transactions on Visualization and Computer Graphics, vol. 12, pp. 1511-1522, Nov. 2006.

[24] TechSmith, Camtasia Studio. Available: http://www.techsmith.com