

Integrating Annotative Notes and Data in the Analysis Process

by

Eric Lee

B.Sc., Simon Fraser University, 2009

Thesis Submitted In Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the

School of Interactive Arts and Technology
Faculty of Communications, Arts and Technology

© Eric Lee 2013

SIMON FRASER UNIVERSITY

Spring 2013

All rights reserved.

However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name: Eric Lee
Degree: Master of Science
Title of Thesis: Integrating Annotative Notes and Data in the Analysis Process

Examining Committee:

Chair:

Dr. Philippe Pasquier
Assistant Professor

Dr. John Dill
Senior Supervisor
Professor Emeritus

Dr. David Darvill
Supervisor
Adjunct Professor

Dr. Chris Shaw
External Examiner
Associate Professor

Date Defended/Approved: 2013 April 03 _____

Partial Copyright Licence



The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website (www.lib.sfu.ca) at <http://summit/sfu.ca> and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author’s written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2011

Ethics Statement



The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

- a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

- b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

- c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

- d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010

Abstract

Analysts need to keep track of their analytic findings, observations, ideas, and hypotheses throughout the analysis process. While some visual analytics tools support such note taking needs, these notes are often represented as objects separate from the data and in a workspace separate from the data visualizations. Representing notes the same way as the data and integrating them with data visualizations can enable analysts to build a more cohesive picture of notes and data. We created a note taking functionality called CZNotes within the visual analytics tool CZSaw for analyzing unstructured text documents. CZNotes are designed to use the same model as the data and can thus be visualized in CZSaw's existing data views.

We conducted a preliminary user study to observe the use of CZNotes and observed that CZNotes has the potential to support progressive analysis, to act as a shortcut to the data, and supports creation of new data relationships.

Keywords: Annotations, Note taking, Visual Analytics, CZSaw, CZNotes, Sensemaking

Acknowledgements

I would like to share my sincere appreciation and gratitude to my supervisor John Dill for his patience, invaluable guidance and encouragement these many years, which made this project successful.

The CZSaw group, comprising John Dill, Rob Woodbury, David Darvill, Chris Shaw, Maureen Stone, Nazanin Kadivar, Dustin Dunsmuir, Victor Chen, Cheryl Qian, Jeffrey Guenther, Saba Alimadadi, and David Botta, had many intense discussions about note taking in analysis and ways to support it in CZSaw. It is impossible to completely unravel the provenance of ideas from these discussions; the best I can do is to offer thanks to all who contributed. Of course, I take full responsibility for any errors and shortcomings in interpreting these discussions as my thesis progressed. Thank you all for your support, guidance and friendships. I also appreciate Richard Arias Hernández for his valuable advice. Thank you to everyone in the SIAT community for this rewarding experience. Wishing you all the best.

I would also like to take this opportunity to thank our sponsors Boeing, AeroInfo Systems, a Boeing Company, and NSERC for making our work possible.

Last but not least, I would like to express my love and gratitude to my mom, Chui Jun Lee, my dad, Ngai Bun Lee, my sisters, Louise, Ashley and Anna, and my grandparents for their unwavering support. To my wife, Penny Ng, for her detailed care and endless love, to my family in Christ for their prayers and spiritual support. To God, my Saviour, I am thankful for His love that gives me peace, joy, and strength to keep striving on until that one fateful day.

Table of Contents

Approval.....	ii
Partial Copyright Licence	iii
Abstract.....	iv
Acknowledgements.....	v
Table of Contents.....	vi
List of Tables.....	ix
List of Figures	xi
1. Introduction	1
2. Related Work	6
2.1. Integrating Notes and Data	10
2.2. Current VA Tool Support for Note Taking	14
2.3. Summary.....	23
3. Design of CZNotes	26
3.1. CZSaw Overview	26
3.2. Design Approach	30
3.3. Integrating <i>CZNotes</i> into <i>CZSaw</i>	34
3.3.1. Entities and CZNotes	34
Entity Extraction Suggestions	35
Creating Relationships Among Entities using CZNotes	36
3.3.2. CZNotes and the CZSaw Script	37
3.3.3. CZNotes and the Dependency Model	38
3.4. Note Recommendations and Progressive Analysis	40
3.5. Collaboration - Sharing CZNotes by Sharing Script Files	41
3.6. Visualizing and Interacting With CZNotes.....	41
A Notes Repository	47
3.7. Interacting with CZNotes.....	48
3.8. Implementation Overview & Software Architecture.....	52

Technology Used	53
Limitations of Our Design	54
3.9. Summary of Design	54
4. User Study	55
4.1. Study Design.....	55
Pre-Study Questionnaire	58
System log of participant’s interactions with CZSaw and CZNotes	58
Discussion Sessions	59
Journaling.....	60
Final Report.....	60
Post-Study Questionnaire	61
4.2. Data Analysis	61
4.3. Limitations	65
5. Results	66
5.1. Questionnaire Results.....	66
5.2. CZNotes Usage Statistics	67
5.3. CZNotes Support for Continued Analysis	68
5.3.1. CZNotes and the Foraging Loop.....	68
5.3.2. CZNotes Support for Continued Analysis.....	71
5.3.3. Reviewing CZNotes.....	75
5.3.4. Building a Story with CZNotes.....	75
5.3.5. Live Links	76
5.4. CZNote Reminders	77
5.5. Using CZNotes to Create New Relationships	79
5.6. Add New Data into CZSaw	81
5.7. Visualizing CZNotes In Data Views	81
5.8. Collaboration and Sharing CZNotes	84
5.8.1. Wiki Versus CZNotes	85
5.8.2. Pursuing Different Threads	86
5.8.3. Live Links and their Need.....	86
5.8.4. Sharing CZNote Scripts.....	87

5.8.5. Using Other Participant’s CZNotes	88
5.8.6. Private Versus Public	89
5.9. CZNotes as a Categorizing Tool	90
5.10. Summary	92
6. Discussion	93
6.1. Limitations	95
7. Conclusion & Future Works	97
7.1. Conclusion	97
7.2. Future Directions	98
7.2.1. CZNotes and Hypertext	100
References	102
Appendices	107
Appendix A. Log File Tags	108
Creating New Notes	108
Adding references to a note	109
Attaching a note to a note	109
How were notes added to a data view?	110
Searching	112
Entity refinements	114
Appendix B. Study Description	115
Appendix C. Pre-Study Questionnaire	118
Appendix D. VA Tool Familiarity Questionnaire	119
Appendix E. Post-Study Questionnaire	120
Appendix F. Participant Consent Form	121
Appendix G. Sample PERL Script used in Data Analysis	123

List of Tables

Table 2-1. Marshall’s note taking types from (Marshall, 1990).....	9
Table 2-2. A comparison of the functionalities of various visual analytic tools.....	24
Table 4-1. Sample of transactions outputted to the CSV file.....	62
Table 4-2. Example rows of data pertaining to the additional CZNote tags outputted to the CSV file.	62
Table 4-3. The initial set of codes based on the initial set of research questions.	64
Table 4-4. The table represents the final set of codes created during our analysis.	65
Table 5-1. The total number of CZNotes created, and edited by all 4 participants.	67
Table 5-2. The number of CZNotes created per participant, the number of times CZNotes have been edited, and the number of times references to entities have been added to a CZNote.....	68
Table 5-3. A count of how many times CZNotes were attached to data in each of data views during the course of the study.....	70
Table 5-4. Number of times data or CZNotes were queried using the get-related function in the Hybrid Graph View. The * represents P1 querying from a CZNote, however, no return type was logged perhaps because CZSaw “froze” and never completed the query.	73
Table 5-5. Number of searches that returned Non-CZNotes only (Any entity type other than CZNotes), CZNotes-Only, and Mixed entity types.	77
Table 5-6. Number of times each participant referenced an entity within a CZNote and number of new relationships created by adding such references to CZNotes that did not exist in the original data set.....	79

Table 5-7. This table shows the frequency of the different ways CZNotes and Non-CZNotes were added to the CZSaw data views: Hybrid Graph View (HGV), Semantic Zoom View (SZV), and Document View (DV). CZNotes were primarily visualized in the Hybrid Graph View. * D & D = Drag and Drop. 82

List of Figures

Figure 2-1. Problem Definition, Visualization, Analysis and Dissemination were identified as four iterative phases of analysis in (Mahyar et al., 2010)'s study. Record-Keeping (i.e. Note taking) was observed to occur throughout all phases of analysis.	8
Figure 2-2. An analysis workspace, involving data (text cards and images) and annotations and notes (post-its) (Robinson, 2008).	10
Figure 2-3. The VA tool HARVEST designed to integrate the synthesis and exploration spaces (Gotz et al., 2006).	12
Figure 2-4. Brushing and Linking was used to connect the note items on the left (the Torch item is selected) to the supporting timeline evidence on the right.	13
Figure 2-5. TRIST (top) and Sandbox (bottom) (Wright et al., 2006).	14
Figure 2-6. A diagramming workspace for creating and organizing notes in Sandbox (Wright et al., 2006).	15
Figure 2-7. The interface for Sense.us with interactive visualizations and a commenting system (Heer et al., 2007). Note: the annotative letters and numbers in the figure are not relevant to this thesis.	18
Figure 2-8. The ARUVI system for scatterplot analysis (top left), note taking and sense making in the Knowledge View (Top right), and analysis state history capture (bottom) (Shrinivasan & Van Wijk, 2008) . Note: the annotative letters and numbers in the figure are not relevant to this thesis.	19
Figure 2-9. Jigsaw's Graph View shows the connection among documents (white nodes) and the entities contained within them (colored nodes) (Gorg et al., 2007).	22

Figure 3-1. CZSaw's Semantic Zoom View. Documents are laid out in a space as nodes. These nodes can be zoomed into with multiple semantic levels. For example, one semantic level contains all the entities contained within the document and a deeper level of zoom reveals the documents text..... 28

Figure 3-2. CZSaw's Hybrid Graph View. The user double clicked on the document “Whitewater, Watergate” to retrieve a set of people nodes (blue nodes) contained in that document. 29

Figure 3-3. CZSaw’s Document View – The left panel is a list of CZSaw data documents, the middle panel lists all the entities contained within the selected document, and the right panel displays the text of the document with the extracted entities highlighted. 29

Figure 3-4. A popup menu to extract a new entity from highlighted text within a CZNote..... 35

Figure 3-5. “Suggested References” feature is used to scan a note’s text for entities to identify and reference. 36

Figure 3-6. The images above show the creation of a relationship between Rob and John through a user created CZNote by identifying the two entities in the same note. 37

Figure 3-7. A screenshot of the script interface containing a list of user actions that were previously recorded. 38

Figure 3-8. A simple dependency graph generated after multiple searches and displaying the search result in a “CZHybridView”. If the user changes the content of the “searchResult3”, the change in content will trigger a propagation to update the contents of “CZHybridView”. 39

Figure 3-9. The Document View is displaying a CZNote written by the user. All CZNotes can be added to the Document View at once using the Green plus button above the first panel on the left. 42

Figure 3-10.	This figure shows how nodes can be represented in various forms. For example, the nodes can represent a single person entity node (blue), an aggregate node containing a set of Reports (white node), or as a list node containing a set of entities.	43
Figure 3-11.	The yellow nodes represent CZNotes containing text content with identified entities e.g. “Billie Codi” and “Albert Webber”.	44
Figure 3-12.	Creating a link from a note to another entity node by dragging the “Add Link” (icon with a chain and green plus) icon from the source CZNote to the destination node.	45
Figure 3-13.	Double clicking on the CZNote (from Figure 3-12) retrieves all entities related to the CZNote.	45
Figure 3-14.	The Semantic Zoom View can visualize CZSaw Documents and CZNotes (represented by the grey rectangles). These notes and documents are clustered based on the entities they contain.	46
Figure 3-15.	The image is of a notes repository containing a list of CZNotes that is searchable and sortable.	47
Figure 3-16.	(a) A blank note can be quick created in the Hybrid Graph View by toggling the “Create Note” icon into the add note mode (b) and then clicking at a point in the Hybrid Graph View to create a CZNote at that location.	48
Figure 3-17.	A new CZNote is being attached to the Report with ID = 12 with the text selected as “explosion that knocked the 747 out of the sky”. The user can also choose to display the CZNote immediately in an existing data view, a new data view or none.	50
Figure 3-18.	A CZNote is being attached to 3 selected entities by dragging the add note icon from one of the selected nodes.	51

Figure 3-19. The architecture of CZSaw and CZNotes. By representing CZNotes in the same way as CZSaw documents, CZNotes is integrated with all the major components of CZSaw. (A) CZNotes can now be visualized in the existing Data Visualizations (B) CZNotes are represented and indexed similarly as the data and thus incorporated into the Dependency Graph Model (C) CZNote interactions are also recorded in the CZSaw Script 53

Figure 4-1. A CZSaw script executed to the transaction t3_note: transactions below the line at t3_note have not yet executed. 62

Figure 5-1. Total number of notes created over the course of study. Green bars: Creation of a new blank CZNote. Orange bars: A CZNote was attached to a CZSaw object (a document or to another entity). (See Chapter 3). 69

Figure 5-2. Attaching a CZNote. (a) P4 highlights text of entity “Bobby Derryberry” in document “Editorial ‘Right Conduct’”. (b) “Attach Note” action generates new CZNote referencing the entity Bobby Derryberry. 71

Figure 5-3. CZNote usage scenario - Opening the CZNote Repository to retrieve 2 existing CZNotes (Yellow nodes) and subsequently retrieve entities referenced within the 2 CZNotes (White nodes). 74

1. Introduction

The amount of information in today's digital world is growing exponentially and analysts must now make sense of increasingly massive amounts of data, especially in fields such as intelligence analysis, business analysis, social media, epidemiology, financial analysis, emergency management and many more. It is thus too complex and time consuming to analyze each individual data item to make sense of it. Tools are needed to help analysts sort through these large data collections to focus on the small subset relevant to their interests. To aid in the battle of analyzing large data sets and capturing and organizing findings to assist in decision-making, analysts can now turn to visual analytical tools. (Thomas & Cook, 2005) define visual analytics (VA) as "the science of analytical reasoning facilitated by visual interactive interfaces". Such tools facilitate data exploration with the aid of salient visualizations, computing science algorithms, statistics, etc. to form visual representations of the data.

In order to build useful VA tools, an understanding of the analysis process is needed. There has been much research done to understand the process analysts use to make sense of data. Analysts iteratively build up a story from the information in the data and their own knowledge, to the best of their ability. (Pirolli & Card, 2005) introduced a notional model for the sense making process. Analysts work in two main loops, the foraging loop, and the *sensemaking* loop. In the foraging loop, analysts explore the data and gather important information into a "shoebox" for collection. When enough information is found, the analyst may proceed to the *sensemaking* loop where they generate hypotheses and attempt to make sense of the data found in the foraging loop. Furthermore, the analyst may iteratively move back and forth between the foraging loop and *sensemaking* loop until the analysis is concluded.

Analysts, as do we all, generate thoughts, ideas, questions, and items to follow up on as they read and analyze data. In this thesis, we refer to such artifacts as notes. Notes may be annotations on the data itself (Marshall, 1990), visualizations, diagrams or informal text and symbols written on a separate notepad (Mahyar, Sarvghad, & Tory, 2010) , or VA tool specific note objects (Gorg, Liu, Parekh, Singhal, & Stasko, 2007; Quinlan, Reed, & Thompson, 2008; Shrinivasan & van Wijk, 2008). Analysts need to record and store such notes in a “shoebox” in the foraging loop or write down hypotheses as a note during the *sensemaking* loop. (Mahyar et al., 2010) research has also shown that analysts create annotations and notes throughout the iterative analysis process. Taking notes is necessary because it is impossible for analysts to keep all of their analytical findings to a complex problem in their memory. (Heuer Richards J., 1999) warns that analysts are more prone to cognitive biases if they primarily do analysis in their heads. Therefore, offloading notes externally can lessen their cognitive load and ensures that such captured artifacts are available, reviewable, and complete. Analysts use notes to support their sense making goals and ultimately come to a conclusion in their analysis. We could argue that, in some cases, such record keeping of notes can be as important, or more so, than the data itself.

Many visual analytic tools now include capabilities to capture such notes, visualizations, hypotheses, evidence, annotations, reminders and to-dos to support data analysis. Some, such as Sandbox for Analysis (Wright, Schroh, Proulx, Skaburskis, & Cort, 2006) and *Entity Workspace* (Bier, Ishak, & Chi, 2006) build links between the note and references to the data. Sense.us (Heer, Viégas, & Wattenberg, 2007) and CommentSpace (Willett, Heer, Hellerstein, & Agrawala, 2011) enable users to annotate and comment on visualizations. Still, other tools enable note taking functionality as an add-on to a web browser such as ScratchPad (Gotz, 2007). Click2Annotate automatically generates an annotation based on a set of templates and a selected subset of data in a scatterplot visualization and can be viewed in a separate annotations window (Y. Chen, Barlowe, & Yang, 2010).

However, most such VA tools represent notes as passive text and as another object to manage during analysis. Furthermore, existing tools expect note taking activities and the organization of notes to be done in a separate workspace within the

tool. Search mechanisms for notes and the data are also often separated. *Sensemaking* requires an integrated knowledge base of “*sensemaking* results” and the data being analyzed to draw a conclusion about the data. Analysis “is a progressive process in which newly synthesized knowledge becomes the foundation for future discovery” says (Gotz, Zhou, & Aggarwal, 2006). However, in many of the current VA tools described above, the analyst must manage the connection between two workspaces in such tools, one for note taking activities and one for data exploration.

Therefore, we propose that VA tools could be designed to provide a greater integration of notes and the data within the system than what current VA tools allow by integrating notes into the same interactive workspace as the data. Marshall observed this natural co-location of annotations and data when analysts worked with pen and paper. She identified three types of note taking styles from her Work Practices Study with professional analysts - annotative notes, interpretive notes, and reminding notes (Marshall, 1990). The annotative notes Marshall described included notes in the margins of a text document and the highlighting of the text, interpretive notes are those that refer to one or more source (or none at all), and reminding notes are those that document a process or list a set of tasks to do. We observed that the final product of collaborative synthesis of Robinson’s study (Robinson, 2008) showed notes mingled with the data. Also, the final products of analysis that analysts create, usually in the form of reports, often include references to the data and the analytical findings. What if, rather than waiting until the production stages of sense making to integrate notes and the data, we allow notes to be integrated with the data throughout all analysis stages including the data foraging and exploration stages, hypothesis generation, and so forth, in a non-disruptive way? Notes should be integrated with the data (visualized in the same visualizations, with notes and data queried in the same way) in order to form a more cohesive picture of one’s progress during all phases of analysis.

One way of integrating notes and data is to simply represent notes in the same way as the data documents. If we represent notes and data as the same type of object as documents (in-terms of implementation) we could potentially see a simplification in the design of VA tools because a single data model can be used to represent notes and data rather than a separate data model for each. Implementation of VA tools would be

simplified as existing data visualizations and interactions with data documents can be applied to notes without significant additional customization efforts to the code. For example, in the VA tool *CZSaw*, *CZSaw's Document View* can be used to view both the data documents and notes (Kadivar et al., 2009). Furthermore, user created notes in a VA tool can be fed back into the knowledge database (now consisting of both notes and data). Such notes then seamlessly appear in future data queries (because notes are returned in the data queries as well as *normal* data documents). This integration of notes and data could also provide a more cohesive knowledge base of both previous analytical findings and the data, and potentially introduce new ways for analysts to create and use notes in VA tools. Our research question is then to study the impact on the analysis process of this integration of notes and data in a VA tool. This also adds the ability to query data and notes simultaneously. In this way, the analyst can continually interact with notes as analysis progresses no matter in the foraging loop or the *sensemaking* loop described by (Pirolli & Card, 2005).

This thesis describes a method we developed for accomplishing this. It shows how this new representation of notes as data documents can make the note taking process more seamless and beneficial to the analyst during the sense making process. Such an integration of notes and data can offer the user new ways to interact with this cohesive knowledge base of data and notes within a VA tool.

We have designed and integrated notes into an existing VA tool focused on analyzing unstructured text, called *CZSaw* (Kadivar et al., 2009). The scope of our research is limited to integrating textual notes into a VA tool focused on the data domain of unstructured text documents. In the context of this research, although we recognize the many types of artifacts that analysts create during the *sensemaking* process as “notes”, we currently focus on only text notes. In the future, notes could contain images and perhaps render “mark-up” languages such as html as well. Currently, most VA tools dealing with unstructured text primarily represent text data as a set of documents with extracted entities (e.g., people, places, events). *Entity Workspace* (Bier, Card, & Bodnar, 2008) and *Jigsaw* (Gorg et al., 2007) are such tools. *CZSaw*, inspired by *Jigsaw's* entity model, provides data visualization views of unstructured text documents; with our approach, this can be applied to notes as well.

We have added to *CZSaw* the capability of creating notes, called *CZNotes*. *CZNotes* were implemented by having them inherit (in the object-oriented programming sense) the properties of a *CZSaw document* (text and extracted entities) and can be visualized in *CZSaw*'s existing data *Document Views*. The *CZSaw* data views consist of a *Hybrid Graph View*, *Semantic Zoom View* and a *Document View*. Furthermore, *CZSaw* captures the semantically meaningful interactions of the user in a script which can be re-played and edited (Kadivar et al., 2009). The interactions with *CZNotes* are also captured in this script.

A preliminary study was run to observe how such an integration of notes and data are used by a group of users (graduate students and a post doctoral fellow) to work on a complex analytical problem. From this study, we observed the participants use *CZNotes* as a container to categorize data and also as a shortcut to access data recorded in *CZNotes*. *CZNotes* were also visualized in existing data views of *CZSaw* and were included in data query results. The *CZNotes* included in the query results proved to be reminders to one participant of the analytical paths taken in past analysis sessions. Furthermore, *CZNotes* were used to create relationships among the data entities, although unintentionally. Furthermore, we observed participants' efforts to share notes with one another even though the system was not designed for easy collaboration with notes. These initial observations provide possible directions for future research regarding the integration of notes and data.

The thesis is organized as follows: Chapter 2 describes related work in note taking and the analysis process, the representation of notes in VA tools, and related VA tools that support note taking. Chapter 3 describes the design of *CZNotes*. Chapter 4 describes a preliminary user study in which participants used *CZNotes* to solve a VAST 2011 mini challenge ("IEEE VAST Challenge 2011," 2011) and how the data was analyzed. Chapter 5 describes the quantitative and qualitative results and chapter 6 presents a discussion of the significance of the results. Chapter 7 concludes with a summary and our future directions for improving the note taking capabilities within an entity based VA tool for text analysis.

2. Related Work

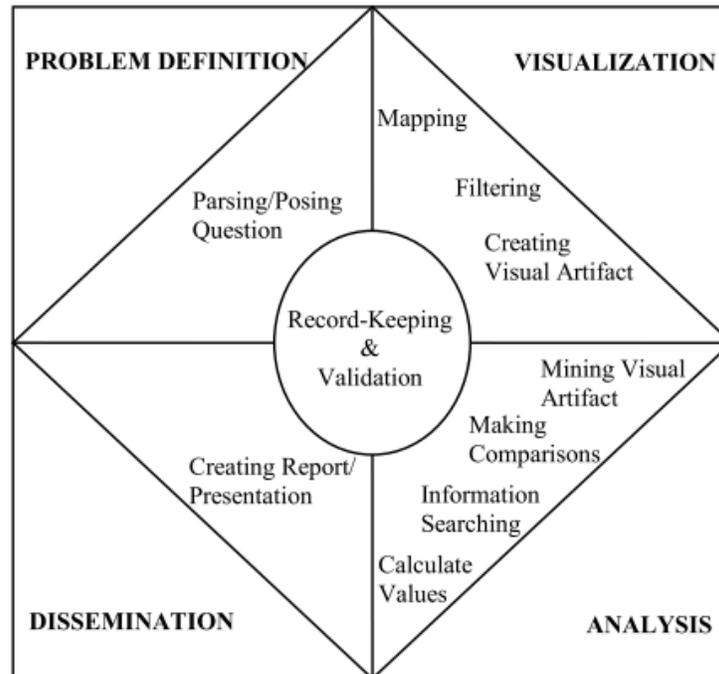
Note taking is prevalent in all areas of life. Today, there are increasing numbers of digital tools available for both specialist and non-specialist to take notes to keep track of items they need to complete, to record important information, and so forth. This is especially important in the context of analysis where analysts often use software to support their analyses. Analysts simply cannot remember everything they read nor all the hypotheses they generate during analysis. (Heuer Richards J., 1999) warns that analysts are more prone to cognitive biases if they primarily do analysis in their head. Offloading notes externally can lessen the cognitive load and thus help ensure that such captured artifacts are available, reviewable, and complete. Such notes are not only for analysts themselves, to remind them of previous analytical findings, but also to share with collaborating team members. Notes can also be the source material to create a summary, presentation or report to disseminate results to decision makers at the conclusion of an analysis.

In this chapter, I review how note taking is central to the analysis process and how current VA tools support note taking. Existing studies have identified different phases during analysis by observing analysts solving complex problems. Such observations indicate that analysts freely move back and forth among the different phases of the analysis process in an iterative fashion and that note taking is often a central activity in this process. (Mahyar et al., 2010; Pirolli & Card, 2005)

(Pirolli & Card, 2005) describe a notional model of sense making, which includes a foraging loop and sense making loop. In the foraging loop, analysts are exploring the data and gathering important information into a “shoebox” for collection. When enough information is found, analysts may move into the sensemaking loop where they generate hypotheses and make sense of the data found in the foraging loop. Note taking can

occur at any of the activities in the above model, especially during the schematizing and hypotheses phases where threads of analysis and associating evidence for and against each hypothesis might be recorded and later reviewed, compared, and shared to determine which “solution” is most viable. This is an iterative model where foraging and sense making are intertwined and thus note taking would be also involved throughout the analysis process.

(Mahyar et al., 2010) conducted a study observing how groups of analysts in a co-located collaborative environment worked on an analysis problem in the business domain. Over the course of each group’s analysis, the investigators found many stages of analysis similar to those documented by other models of analysis (Isenberg, Tang, & Carpendale, 2008; Pirolli & Card, 2005). In addition, (Mahyar et al., 2010)'s framework for collaborative analysis derived from their study (Figure 2-11) identified that note taking (or record-keeping as they termed it) occurred throughout the analysis process. The participants created notes for themselves and for the group in the form of written notes on paper and screen captures of visualizations from the VA tool used. The researchers found multiple uses for such notes for facilitating problem solving, for example, “by recording the direction and sequence of the steps taken” (Mahyar et al., 2010), creating reminders, and eventually the final report.



Reprinted from Mahyar, N., Sarvghad, A., & Tory, M. (2010). A closer look at note taking in the co-located collaborative visual analytics process. *IEEE VAST*, 171–178. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5652879 © 2010 IEEE. Used with permission.

Figure 2-1. Problem Definition, Visualization, Analysis and Dissemination were identified as four iterative phases of analysis in (Mahyar et al., 2010)’s study. Record-Keeping (i.e. Note taking) was observed to occur throughout all phases of analysis.

In an earlier work practice study, (Marshall, 1990) identified an iterative analysis process where analysts made annotations on paper and electronically. She worked with two groups of analysts, where one group analyzed primarily text data while another analyzed images and video. Through observations she noticed that analysis appeared to be an iterative process. “[N]otetaking can be driven by information gathering, or it can be driven by the production of a written product”, says (Marshall, 1990). She describes, “[t]he analyst may [mark up a hardcopy of an informational email] (a notetaking activity) and file it, or it may fuel a later writing process. Activities in the interpretation process, reading, notetaking, and filing may suggest additional retrieval needs as the analyst tries to make sense of the information at hand” (Marshall, 1990).

Marshall also describes the analysts' need to file notes, observing that analysts file notes and the source material together (i.e, notes and data together) but the methods used varied among analysts. Most filing strategies were a combination of storing items chronologically and by category or theme of the notes and documents, and the choice of strategy may be influenced by the phase of the analytic process as well (**Error! Reference source not found.**). These filed notes and documents will be retrieved when they need to be shared and/or when writing a final report. Marshall states that, "most of the analysts pull out their collection of materials on the desired subject to create a context for writing" (Marshall, 1990). Regardless of the type of data (text versus video and images) being analyzed, in both cases, Marshall observed the same three types of note taking styles - Annotative notes, interpretive notes, and reminding notes (Table 2-1).

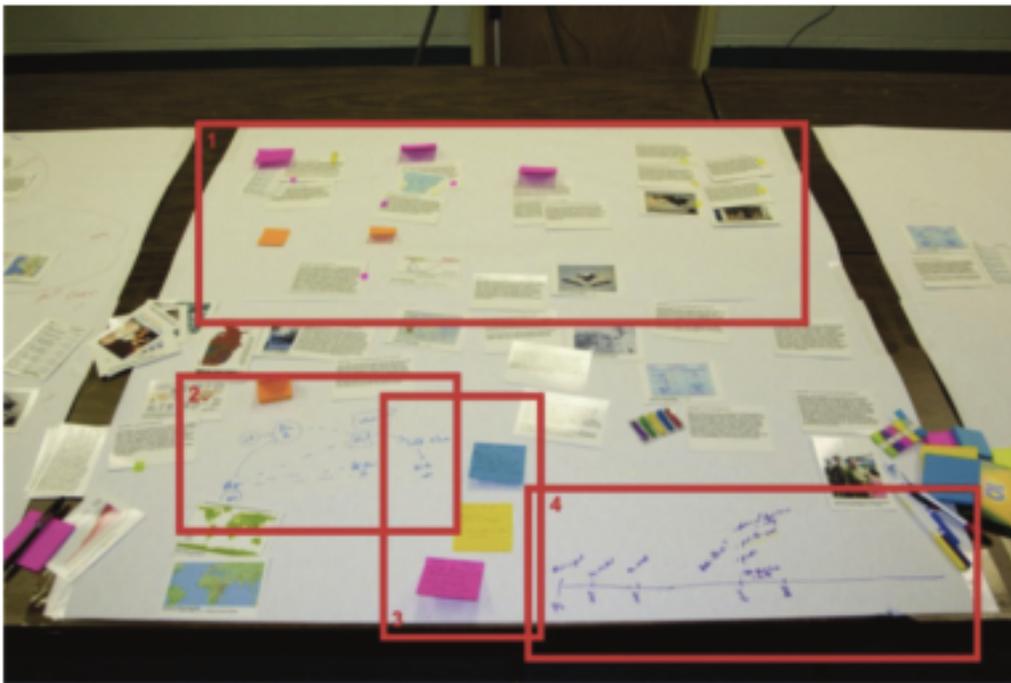
Annotative Notes	Interpretive Notes	Reminding Notes
Text highlighting (keywords, and regions)	Interpretive or integrative notes referring to one or more source	Auxiliary notes documenting a systematic process
Annotations and marginalia	Text notes note referring to any source directly	Auxiliary notes listing an agenda of subtasks

Table 2-1. Marshall's note taking types from (Marshall, 1990).

Annotative notes are mark-ups on the hard copy of the data documents, which could be highlighting the text or annotations and marginalia on the document. Interpretive notes are written text "to record conclusions they have reached or material they have integrated from several sources"(Marshall, 1990). Such interpretive notes were sometimes written in an on-line text editor. Reminding notes serve as a means "to jog [an analyst's] memory about things to do or portions of procedures to follow." Such notes are often written on sticky notes. Marshall speculates that reminding notes may be a way to preserve procedural knowledge of how analysis was done.

2.1. Integrating Notes and Data

Analysis requires the continual build up of understanding of the data. This is referred to as *synthesized knowledge* by (Gotz et al., 2006) and is usually represented by notes of some sort. A melding of notes and data would be a logical approach to take when designing a VA tool. Accordingly, the main design goal of *CZNotes* is the integration of notes and the data documents within the text analysis tool, *CZSaw*.



Reprinted from Robinson, A. C. A. (2008). Collaborative synthesis of visual analytic results. *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on* (pp. 67–74). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4677358. © 2008 IEEE. Used with permission.

Figure 2-2. An analysis workspace, involving data (text cards and images) and annotations and notes (post-its) (Robinson, 2008).

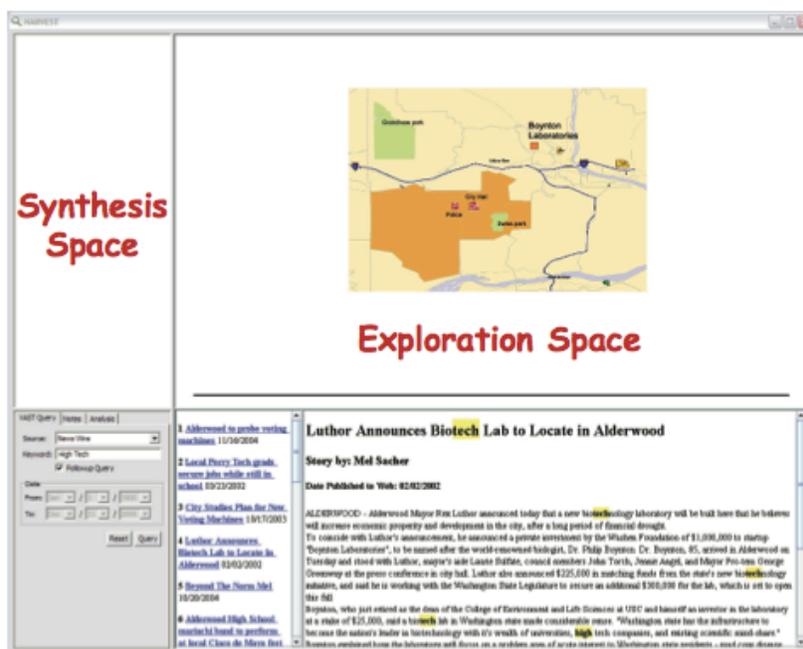
(Robinson, 2008) studied analysts working in a co-located collaborative environment. Participants collaborated in pairs with a physical blank workspace, physical artifacts (the data) and markers or sticky notes for notes and annotations to make sense of the data. The resulting workspaces varied in organization metaphors with the artifacts (the data),

sticky notes and annotations with markers mixed together (Figure 2-2). As the analysis session progressed, usually one analyst would take on the task of recording hypotheses, key facts and other pertinent information. In the end, analysts wrote hypotheses on sticky notes and placed them on the workspace. They also formed hypothesis groups and timelines. Analysts ultimately combined their synthesized knowledge and the data in the same workspace to support continued sense making. We observed that the final product of the participants' analyses in this study shows that the integration of the data and the synthesized knowledge in the same workspace.

(Mahyar et al., 2010) observed participants using *SAP Business Objects Explorer* in co-located collaborative sense making tasks. They suggested that VA tools could integrate notes as part of the history mechanism as well as provide a notebook style workspace for collecting findings from multiple sources or for creating "reminding" notes. They suggest notes should be linked to the data in the system that the note references. Furthermore, they say, "[notes] could be treated by the system as if they were artifacts such as charts. For example, they could potentially be added to a chronological history in the same way that a chart would be added" (Mahyar et al., 2010). This suggests elevating the note object to be a higher (first) class object.

Some VA tools have already taken on the task of providing a more integrated workspace for combining the tasks of foraging - or information retrieval - and sense making into a fluid system. We briefly review the main work in this area.

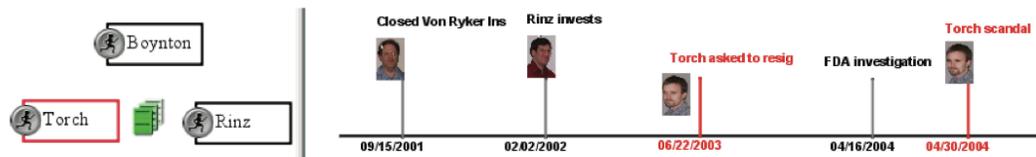
(Gotz et al., 2006) described the need for the examination of existing understanding of the data and the synthesis of new analytic knowledge together. Analysis "is a progressive process in which newly synthesized knowledge becomes the foundation for future discovery." Their system, HARVEST, allows the user to both annotate existing information and refine existing annotations. HARVEST supports the "visual exploration of both existing information and synthesized knowledge. This is achieved by dynamically incorporating synthesized knowledge together with raw information to form a single coherent visual exploration context" (Gotz et al., 2006).



Reprinted from Gotz, D., Zhou, M., & Aggarwal, V. (2006). *Interactive Visual Synthesis of Analytic Knowledge*. 2006 IEEE Symposium On Visual Analytics And Technology (pp. 51–58). IEEE. doi:10.1109/VAST.2006.261430. © 2006 IEEE. Used with permission.

Figure 2-3. The VA tool HARVEST designed to integrate the synthesis and exploration spaces (Gotz et al., 2006).

(Gotz et al., 2006) explains that one of their goals in their analysis tool, HARVEST, is to link the synthesized knowledge with the evidence in the exploration space. In their tool, the synthesis space is placed next to the exploration space (Figure 2-3). These spaces are linked such that if the user clicks on the hypothesized suspect, for example, “Torch” in Figure 2-4, the corresponding evidence linked to the suspect is also highlighted in the exploration space. They state, “... it is important to allow users to quickly check the correspondence of the synthesized knowledge and its evidence at any given point of the investigation.”

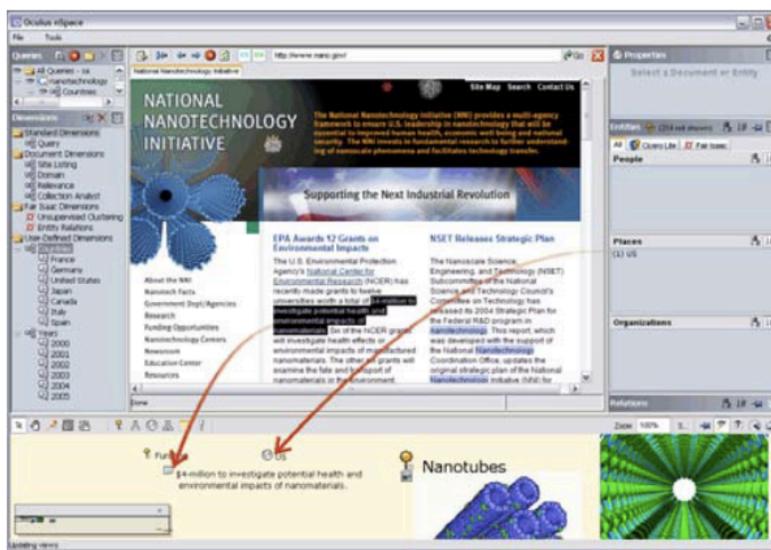


Reprinted from Gotz, D., Zhou, M., & Aggarwal, V. (2006). *Interactive Visual Synthesis of Analytic Knowledge*. 2006 IEEE Symposium On Visual Analytics And Technology (pp. 51–58). IEEE. doi:10.1109/VAST.2006.261430. © 2006 IEEE. Used with permission.

Figure 2-4. Brushing and Linking was used to connect the note items on the left (the Torch item is selected) to the supporting timeline evidence on the right.

Another research effort supporting a tighter integration of information foraging and sense making (using notes) is *nSpace* (Canfield & Sheffield, 2011). *nSpace* is a combination of two tools: *TRIST* (Jonker, Wright, Schroh, Proulx, & Cort, 2005), an Information Retrieval tool and *Sandbox* (Wright et al., 2006), a sense making and evidence marshalling tool.

A goal of *nSpace* is to support “fluid iterative discovery, blurring the boundaries between information retrieval and analysis” by combining *TRIST* and *Sandbox* (Wright et al., 2006). Although both of these components are within the same system, there is still a spatial disconnection between the information retrieval workspace and the analysis workspace (Figure 2-5).



Reprinted from Wright, W., Schroh, D., Proulx, P., Skaburskis, A., & Cort, B. (2006). *The Sandbox for analysis: concepts and methods. Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 801–810). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1124890>. © 2006 ACM, Inc. Used with permission.

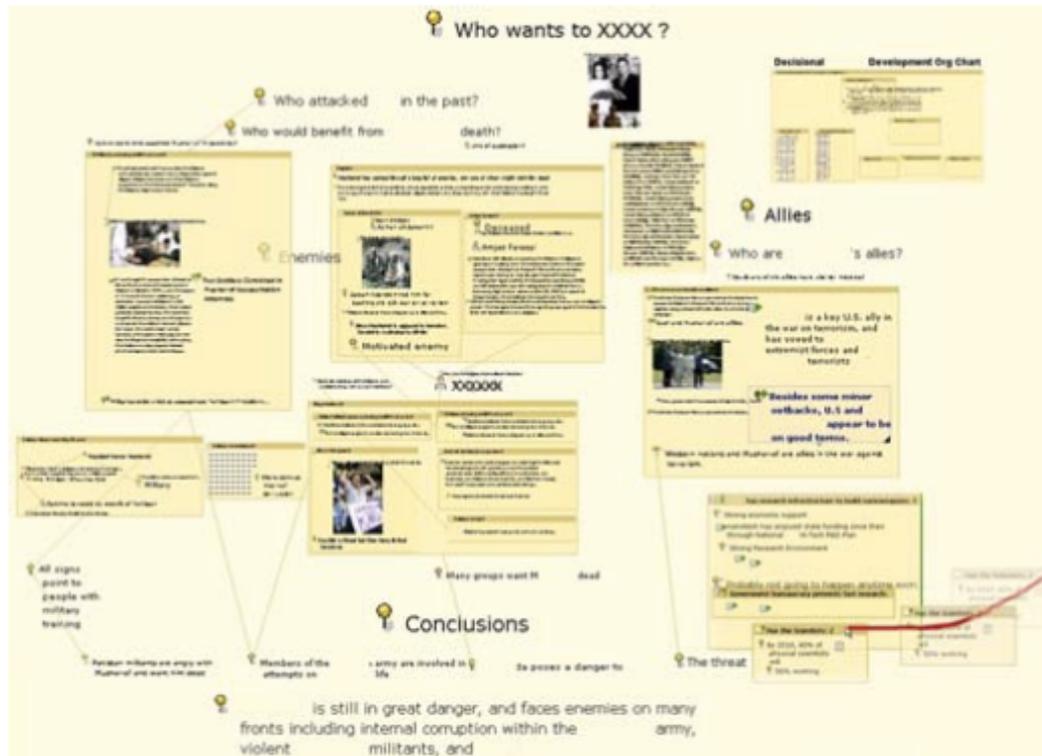
Figure 2-5. TRIST (top) and Sandbox (bottom) (Wright et al., 2006).

Entity Workspace (Bier et al., 2006) also tries to achieve support for the foraging and sense making loops by incorporating a CorpusView document reader (foraging) and the *Entity Workspace* (sensemaking), an evidence panel, into a single application.

Although *HARVEST* (Figure 2-3), *Entity Workspace* and *nSpace* (Figure 2-5) attempt to integrate the synthesis space and the exploration space into the same tool the two functions are still separated from one another as separate panels within a window. We attempt to go a step further with the design and development of *CZNotes* and merge the “Synthesis Space” with the “Exploration Space” into a single space (Figure 2-3). We examine the possible benefits of this design to analysts at the end of this chapter.

2.2. Current VA Tool Support for Note Taking

We now consider other aspects of note taking in existing VA tools. Note taking should be considered when designing any VA tool, no matter which data domain the tool focuses on (e.g. multi-dimensional data or unstructured text data).



Reprinted from Wright, W., Schroh, D., Proulx, P., Skaburskis, A., & Cort, B. (2006). *The Sandbox for analysis: concepts and methods. Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 801–810). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1124890>. © 2006 ACM, Inc. Used with permission.

Figure 2-6. A diagramming workspace for creating and organizing notes in Sandbox (Wright et al., 2006).

In our review of existing VA tools, we focus on those tools that primarily support text analysis since CZNotes was developed and integrated within a text analysis tool, CZSaw. We look at representations of notes, interactions with notes, the relations of notes and entities, note recommendations and collaboration using notes. We compare CZNotes to these tools and also provide a high level description of CZNotes.

There are VA tools that allow the user to create notes and represent them as text objects, comments or nodes (like post-its). These can be manipulated in a free-form organizational space like nSpace's Sandbox (Wright et al., 2006). See Figure 2-6.

Tools such as Sandbox, *Jigsaw*, nSpace and *Entity Workspace*, represent notes as text objects and enables the user to spatially arrange these notes in a meaningful manner, compare findings, etc. These tools often include the ability to draw links among the notes or create links between notes and the data sources to indicate relationships.

With nSpace, a user can drag data retrieved from *TRIST* (exploration space) into Sandbox (the synthesis space) to create a new note artifact with a linkage to the data source that is automatically maintained (Wright et al., 2006).

Entity Workspace provides a *Document Viewer* and search capabilities in a tool called CorpusView. Next to CorpusView is an Evidence Panel, a notebook like workspace where user defined notes and data entities can be “snapped” together to create structured statements or groupings (Bier et al., 2006). The analyst can then arrange these statements and groups spatially in a meaningful way. *Entity Workspace* also provides a feature for users to quickly generate a note in the workspace by clicking on entities in the CorpusView. This feature copies a phrase from a data document with entity extraction (entities and entity extraction is reviewed below) into the *Entity Workspace* notebook as a new item (Bier et al., 2006).

IN-SPIRE is a text analysis tool which can analyze and group textual data documents into clusters using its Galaxy view or ThemeView tools (Pacific National Northwest Laboratory, 2013; Wong et al., 2004). The user can place documents into groups. These groups can each represent a different hypothesis. IN-SPIRE enables the user to mark a document as supporting or refuting a hypothesis (group) and visualizes this information in a table for further analysis.

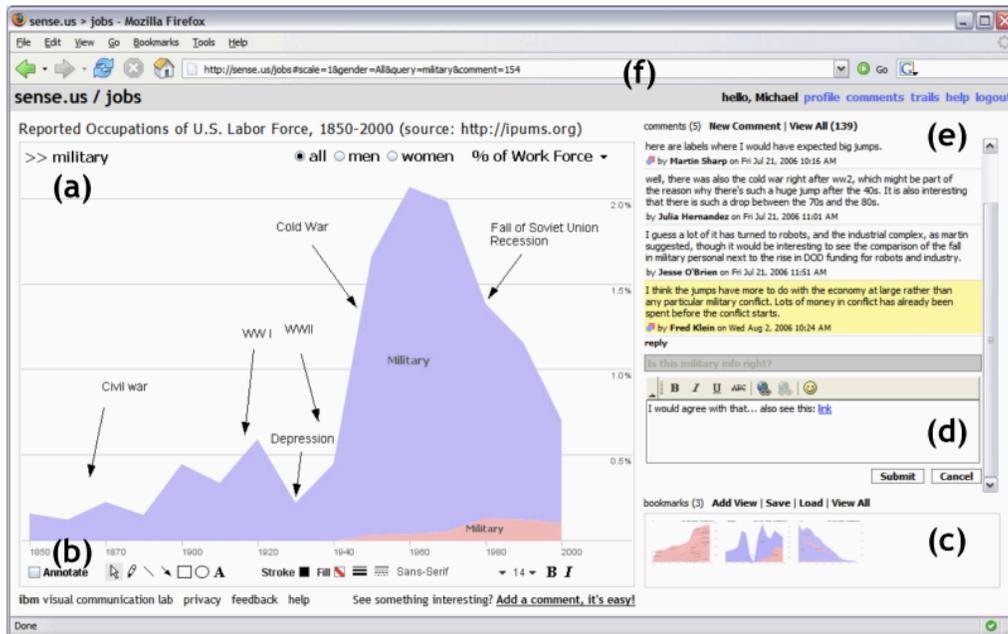
Touch2Annotate (Y. Chen, Yang, Barlowe, & Jeong, 2010) and Click2Annotate (Y. Chen, Barlowe, et al., 2010) also provide an automatic generation of notes in natural language style in the domain of multi-dimensional data. Although this semi-automatic approach is beneficial in collecting facts and phrases about the data quickly, free-form user-defined notes are still required to allow the analyst to record hypotheses, ideas, reminding notes, etc., which cannot be generated automatically. Thus, the ability to create free-form notes is available in both *Entity Workspace* and Touch2Annotate and

Click2Annotate. These annotations can be viewed and edited in a separate annotation window.

Jigsaw provides a Tablet view that allows the user to add both data entities and notes into the same view. Linkages can be drawn amongst the items and it also supports the creation of a timeline of events to support “story building” (Liu et al., 2010).

When we designed *CZNotes*, we intentionally did not build new data views into *CZSaw* to represent and interact with notes but rather utilized existing data views to interact with *CZNotes*. This was possible because our design uses the same data model to represent notes (i.e. *CZNotes*) as the *CZSaw* documents (text document with extracted entities). With minimal additions to the actual code, we can visualize *CZNotes* in the existing *CZSaw* data views. *CZSaw’s Hybrid Graph View* provides a diagramming capability for interacting with *CZNotes* similar to the above tools without creating a *CZNote* specific view. The user can create new (blank) notes in the *Hybrid Graph View* as well as create notes referencing a data document or entity in any of the existing data views. The key result is that *any CZSaw* data view (i.e. the *Document View*, *Hybrid Graph View*, and *Semantic Zoom View*) can visualize and offer interactivity with *CZNotes*. We describe these aspects further in chapter 3.

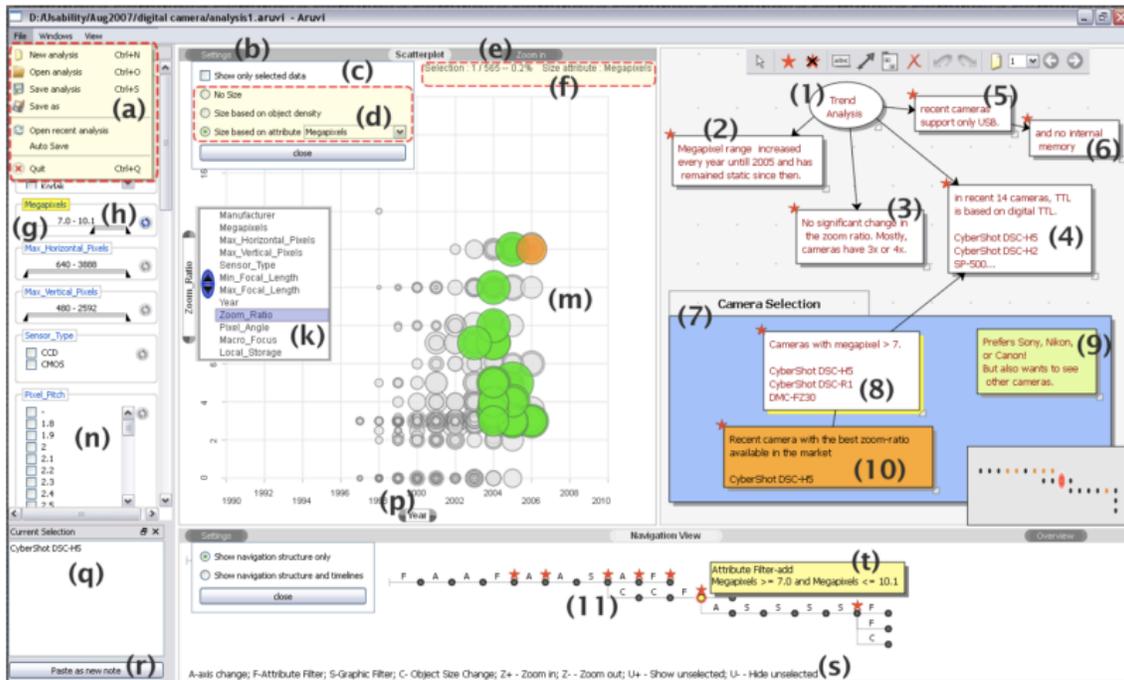
Several VA tools capture the states of the analysis process for provenance purposes and some allow the user to annotate these analysis or visualization states. For example, both Sense.us (Heer et al., 2007) and CommentSpace (Willett et al., 2011) provide a collaborative setting to perform visual analytical tasks and also write comments associated with visualization states. Sense.us “... supports view sharing [of multiple visualization types], discussion, graphical annotation, and social navigation ...” (Heer et al., 2007), allowing the user to annotate and attach comments to visualization states (Figure 2-7). Users can browse and search a global listing of these comments in an asynchronous collaborative fashion and access visualizations through the linkage of a comment. Users could also serendipitously see other user comments if they happen to “land” on a visualization state about which another user had previously commented.



Reprinted from Heer, J., Viégas, F. B., & Wattenberg, M. (2007). *Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization. Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (p. 1029). New York, New York, USA: ACM Press. doi:10.1145/1240624.1240781. © 2007 ACM, Inc. Used with permission.

Figure 2-7. The interface for Sense.us with interactive visualizations and a commenting system (Heer et al., 2007). Note: the annotative letters and numbers in the figure are not relevant to this thesis.

CommentSpace is similar to Sense.us in that it provides a collaborative visual analysis environment and the ability to add comments to visualization states but in addition, it includes the ability to add pre-defined tags “hypothesis”, “Question”, and “To-do” to the comments to support a more accurate and consistent categorization of comments. Additionally, comments may be linked to multiple visualization states which was not possible in Sense.us (Willett et al., 2011).



Reprinted from Shrinivasan, Y. B., & Van Wijk, J. J. (2008). Supporting the analytical reasoning process in information visualization. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08* (pp. 1237–1246). New York, New York, USA: ACM Press. doi:10.1145/1357054.1357247. © 2008 ACM, Inc. Used with permission.

Figure 2-8. The ARUVI system for scatterplot analysis (top left), note taking and sense making in the Knowledge View (Top right), and analysis state history capture (bottom) (Shrinivasan & Van Wijk, 2008) . Note: the annotative letters and numbers in the figure are not relevant to this thesis.

ARUVI provides scatterplot visualizations in combination with what they call a “knowledge view” diagramming workspace for recording notes similar to Sandbox. In addition, ARUVI records analysis states (as shown in the bottom panel of Figure 2-8). ARUVI also allows users to link a note to a visualization state, indicating such links with a star (Figure 2-8). Clicking a starred entity revisits the linked visualization state (Shrinivasan & Van Wijk, 2008).

Analysis requires a continual build-up of the analyst’s own discoveries during analyses over time to determine the underlying story embedded in the data. This requires the analyst to build connections among previously created analytical findings

(notes) and new ones. This becomes increasingly difficult as an analysis progresses however, because our memories are limited. That is why some VA tools, such as Analytic Stream (Nobarany, Haraty, & Fisher, 2012), InsightFinder (Cheng & Gotz, 2009), and Scratchpad (Gotz, 2007), include a “note recommendations” feature to remind users of previously created notes. The idea behind note recommendations is to non-disruptively display past notes to the user to remind them of previously synthesized knowledge relevant to the current analysis context. These recommendations can be used as a memory aid or to potentially spark “aha” moments to see a connection between a past finding and a new data document, for example. This can also support collaboration by recommending notes created by collaborators during an analyst’s analysis session.

One such tool is Analytic Stream (Nobarany et al., 2012). Analytic Stream is a text analysis tool that provides a document reader and note taking capabilities. Notes can be created and shared amongst collaborators. The method to determine what artifacts (“notes”) to recommend to the analyst is based on content similarity between artifacts and the data document currently being viewed by the user. The recommended notes show up in a non-disruptive side panel of the system sorted by relevance.

InsightFinder is a web based tool to support the post-search activities to find specific content within a resulting web page (Cheng & Gotz, 2009). First, users can take notes within the browser using InsightFinder during their web searching. As the search goes on, past notes created by the user are compared with the user’s current web page view. If there is some connection between certain sections of a web page, “page-unit”, to a previously created note, that page-unit will be recommended to the user in a priority sorted list. The goal is to make a connection between a user’s previous notes and the new content they are viewing.

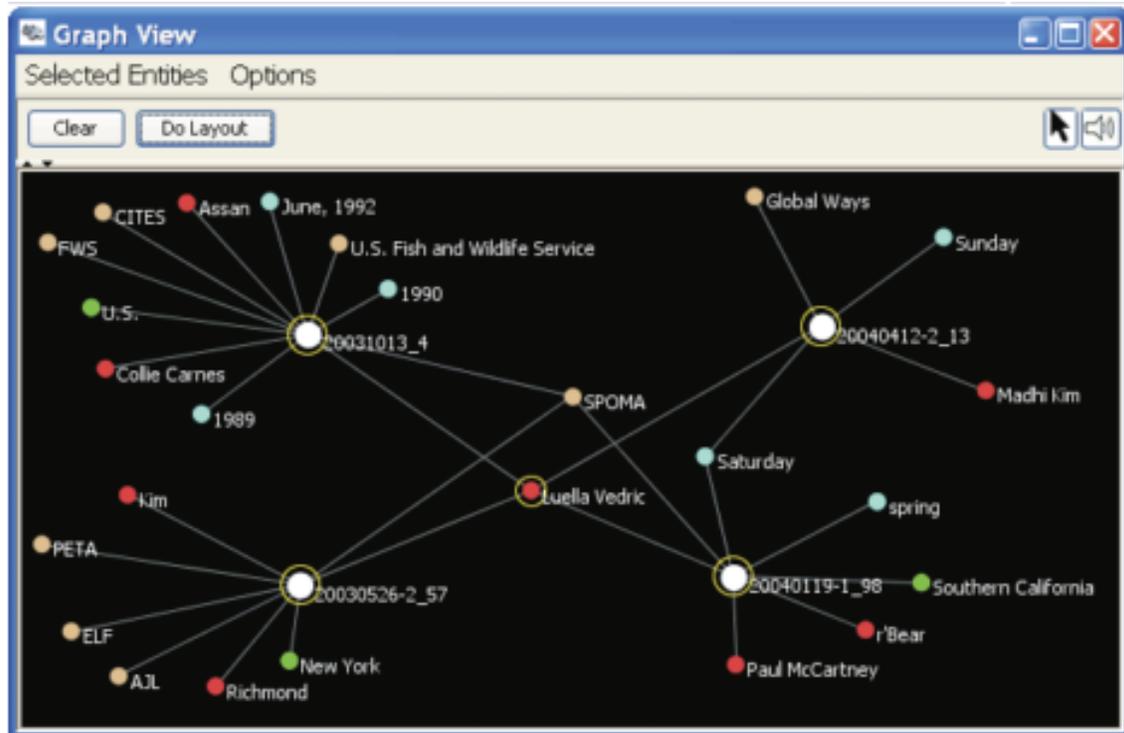
In a similar tool called ScratchPad, notes can be created in a side panel, a “notes sidebar”, of a web browser. These notes are compared with the current web page in view. If there are any content similarities or relevance among the notes and the webpage then the notes in the “notes sidebar” will be highlighted to reveal a connection between the note and the web page (Gotz, 2007).

Entity Workspace (EW) provides more comprehensive recommendation system not only for recommending notes but also recommending relevant data documents to review (Bier, Card, & Bodnar, 2010). EW focuses on making entities (i.e. persons, organizations, places, etc.) a first class object in the system. Unlike previously described tools, the basis for recommendations in EW is an entities' graph that is constructed as the analyst combines entities into an "evidence group" or creates direct relationships between entities. The entity graphs generated from 2 entity graphs, potentially from 2 different analysts, can be combined. This combination can generate new connections that can be used to inform analysts of related entities from other analyst's notebooks (Bier et al., 2010). We describe the role of entities in some VA tools and entities and note taking later in this chapter.

CZNotes takes a novel approach to providing recommendations by simply including *CZNotes* as search results in the *CZSaw* data query results (since the main goal of *CZNotes* is to integrate them with the data). We hypothesize that when the user sees the *CZNotes* in the results in search results and also in implicit queries, for example double clicking a node in the *Hybrid Graph View* to retrieve data, could potentially remind analysts of their previously created *CZNotes*. *CZNotes*, however, does not provide peripheral note recommendations based on the context. These aspects are further described in chapter 3.

Another note property that can be inherited from the data document representation in most unstructured text analysis tools is the idea of entity extraction. In the domain of unstructured text data analysis, tools such as *TRIST* (Jonker et al., 2005), *Jigsaw* (Stasko, Görg, & Spence, 2008), *CZSaw* (Kadivar et al., 2009) and *Entity Workspace* (Bier et al., 2010) have employed named entity recognition tools to extract some structure from text data. (Bier et al., 2010) makes the point that analysts think in terms of entities, i.e. the people, places, things in the data (primarily unstructured text data). Interactions with entities rather plain text makes the analysis more structured and enables the user to do more with the data such as building entity graphs in the case of *Entity Workspace* and working with entity objects rather than plain text. *TRIST* can utilize

a variety of entity extraction systems to parse the text data for entities such as people, places, organizations and display them in an interactive panel (Jonker et al., 2005).



Reprinted from Gorg, C., Liu, Z., Parekh, N., Singhal, K., & Stasko, J. (2007). *Visual analytics with jigsaw*. 2007 IEEE Symposium on Visual Analytics Science and Technology (pp. 201–202). Ieee. doi:10.1109/VAST.2007.4389017. © 2010 IEEE. Used with permission.

Figure 2-9. Jigsaw's Graph View shows the connection among documents (white nodes) and the entities contained within them (colored nodes) (Gorg et al., 2007).

Jigsaw also includes a named entity recognition process at the time of data ingestion (Stasko et al., 2008). This process will identify entities such as people, places, organizations, etc. By identifying entities in Jigsaw, the analyst can navigate in a graph view among all the documents containing a selected entity and, vice versa, find all entities found in a specific document (Figure 2-9).

CZSaw, an entity based text analysis tool, was inspired by Jigsaw and similarly uses an entity-structured approach (Kadivar et al., 2009). However, these named entity recognition tools are far from perfect, resulting in missed or misidentified entities. Thus

VA tools like *Jigsaw* and *CZSaw* provide the user with the ability to refine the entity extraction in real-time. For example, the user can remove wrongly identified entities and extract new entities within the data documents.

By representing *CZNotes* the same way as the data documents within *CZSaw*, *CZNotes* inherit such entity extraction capabilities and enables the analyst to identify entities within a *CZNote*. This entity structure is also a means to create relationships among documents. For example, two entities are related if they are both contained in the same document. Therefore, *CZNotes* could be used to create relationships among entities if, for example, an analyst extracts two entities within one *CZNote*, than these two entities are related by being contained in the same *CZNote* – bibliographic coupling.

Many VA tools such as *CommentSpace*, *AnalyticStream*, and *Entity Workspace* provide support for collaboration for analysis and sense making tasks. Though *CZSaw* and the design of *CZNotes* did not fully take collaboration into consideration, there is still an avenue within *CZSaw* in which *CZNotes* can be shared. *CZSaw* was used to solve the VAST 2010 challenge and team members shared copies of their *CZSaw* script, containing their analysis actions, which can be replayed on another analyst's local machine. As *CZNotes* will also be recorded in the *CZSaw* script, sharing of the script can facilitate the sharing of *CZNotes*. We describe the methods of possible collaboration in the design chapter – Chapter 3.

2.3. Summary

Table 2-2 provides a comparison of some of the VA tools we've reviewed based on entity usage, the type of data analysis, how notes are represented and interacted with, and whether the tool recommends past notes to the user during analysis.

VA Tool	Entity Based	Data Analysis Type	Note Format and Interaction	Note Recommendations
<i>CZNotes</i>	Yes	Text	Text, CZSaw Data Views	Return of notes in subsequent query results
Analytic Stream	Yes	Text	Text Diagramming	Notes recommended based on data view content
ARUVI	No	Multivariate Scatterplots	Text, Diagramming	Visual indicator of note made at a historical analysis state
Click2Annotate / Touch2Annotate	No	Multivariate Scatterplots	Text	No
CommentSpace	No	Visualization States	Text Comments	Tags and Links attached to Comments
<i>Entity Workspace</i>	Yes	Text	Text, Notebook Style	Yes - using entity graphs
InsightFinder	No	Text, Webpages	Text	Notes recommended based on data view content
IN-SPIRE	No	Text	Enables grouping documents into hypothesis categories	No recommendation for notes, but has tools to calculate the correlation among documents
Jigsaw	Yes	Text	Text, Diagramming	No
ScratchPad	No	Text, Webpages	Text, Diagramming	Notes recommended based on data view content
Sense.us	No	Multi-dimensional	Commenting	No
Tableau	No	Multi Dimensional	Scatterplot, Geoviz	No
<i>TRIST & Sandbox</i>	Yes	Text	Text, Diagramming	No

Table 2-2. A comparison of the functionalities of various visual analytic tools.

Based on our review of the related works, we draw suggestions for design requirements for *CZNotes*; we describe the design in the following chapter. The main concept behind *CZNotes* is to integrate analysts' notes (i.e., insight, evidence, hypotheses, etc.) with the data they are analyzing. We hypothesize that such integration will have the following benefits:

1. "Natural" recommendations of existing notes by returning *CZNotes* as search results of the data queries.
2. *CZNotes* can support the construction of new connections among entities and documents.
3. Simpler implementation and support of note taking in existing VA tools by using a single data model for *CZSaw* documents and for *CZNotes*.
4. New affordances of how analysts use notes by allowing interactions between notes and the data in the same, existing, *CZSaw* data views.

3. Design of CZNotes

In this chapter we briefly review the capabilities of the *CZSaw* visual analytics system and then describe the design of the *CZNotes* note taking system for *CZSaw*, including a description of how we incorporated *CZNotes* into the *CZSaw* system.

3.1. CZSaw Overview

CZSaw is a Visual Analytic tool, the first version of which was developed by the Visual Analytics Research Group at Simon Fraser University in 2008. *CZSaw*'s capabilities were intended to provide interactive visualizations for analyzing collections of unstructured text data, such as news articles, blog postings, intelligence reports, and so forth. A key goal of *CZSaw* is to support the capture and analysis of the analysis process itself.

Many VA text analysis tools represent text data as a set of documents with their extracted entities, e.g., people, places, events. *CZSaw* is such a tool, whose original development was inspired by *Jigsaw* (Stasko et al., 2008). There are many benefits to extracting entities from unstructured text. Primary among these is the structure such entity extraction brings to the data. Additionally, connections can be built among entities and the text documents to enable queries.

To extract entities for use in *CZSaw*, text documents are processed using a Named Entity Recognizer (NER) library to identify common entities within the text. The *CZSaw* group worked with the Simon Fraser University Natural Language Processing (SFU NLP) research group to integrate NER capability to tag entities within the unstructured text. We will refer to these text documents with extracted entities a *CZSaw* document.

Each entity has a type and various attributes that is pre-defined, but may be user definable in future versions of *CZSaw*. A *CZSaw* document is also an entity of type Report. For example, a news article would be represented in *CZSaw* as a Report entity with an attribute containing the text of the article and another attribute listing all the entities contained within it (e.g., people, locations, dates, etc). Data visualizations and data queries all act upon these fundamental Entity objects.

A limitation for entity-based systems is that NER cannot yet determine if two entities with same string are the same entities or if they should be distinct entities – distinct IDs. *CZSaw* initially treats all entities of the same name as different entities. It is up to the user to use *CZSaw*'s entity merging feature to combine entities together – to have the same entity ID.

With this entity structure, we utilize connections among entities and the documents to support queries and interactive analysis. When documents contain the same entity, they are connected through the common entity and have a “bibliographic coupling” relation. For example, if documents A and B contain the person entity “Bob”, then documents A and B are related by bibliographic coupling.

NER tools are not perfect, however, and may miss entities, or mis-identify them. To alleviate this problem, *CZSaw* incorporates a capability for users to refine the entities extracted by the NER tool by manually selecting text within documents and tagging them as entities, remove existing entities or merging entities.

CZSaw provides 3 data views for exploring the data: the *Semantic Zoom View*, the *Hybrid Graph View*, and the *Document View*.

The *Semantic Zoom View* (SZV) is focused on visualizing subsets of the documents. The SZV allows the user to apply clustering algorithms to visually group (cluster) documents with similar content. The SZV also allows the user to drill down (or up) different semantic levels of a CZSaw document. For example, one level consists of a listing of all entities within a document. Another level contains the full (scrollable) text of the document (Figure 3-1).

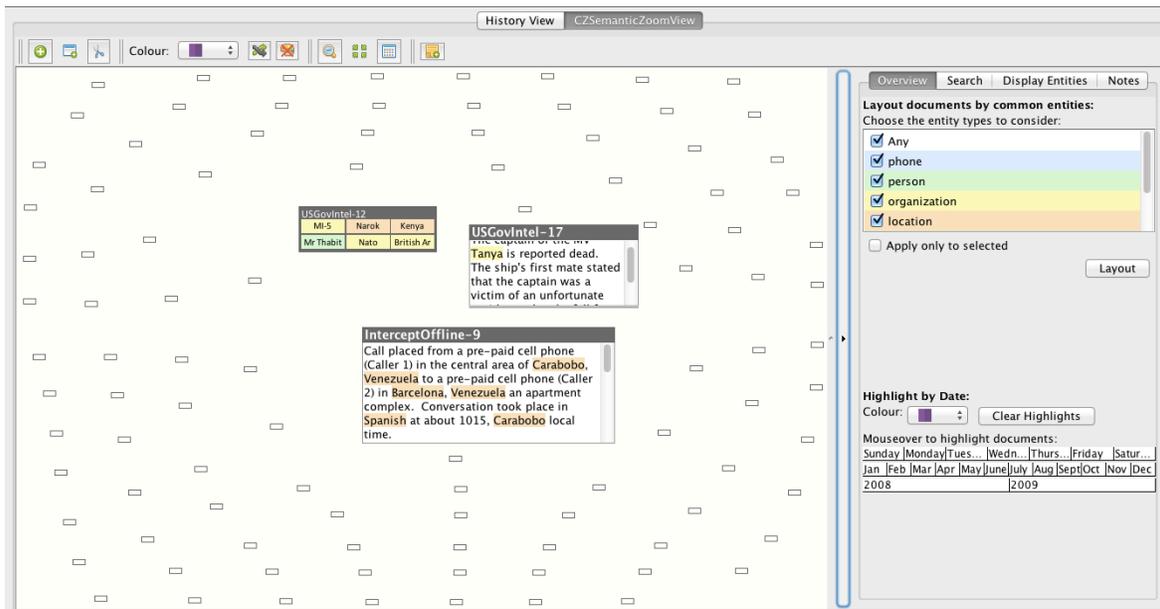


Figure 3-1. CZSaw's Semantic Zoom View. Documents are laid out in a space as nodes. These nodes can be zoomed into with multiple semantic levels. For example, one semantic level contains all the entities contained within the document and a deeper level of zoom reveals the documents text.

The *Hybrid Graph View* (HGV) visualizes entities and documents in a node and link graph style view. This graph view allows the relationship between documents and entities to be shown. Double clicking on a node will execute a query to retrieve related entities and documents and display a connection to it in the graph (Figure 3-2). Thus, the HGV enables the user to discover and explore entity relationships within the data.

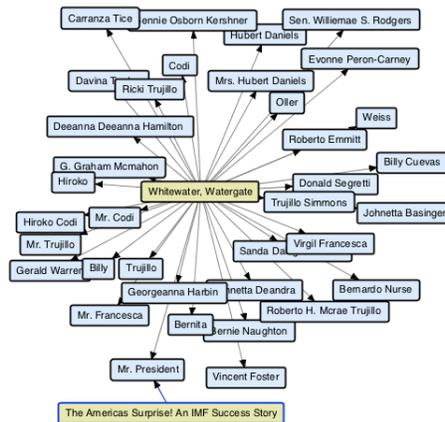


Figure 3-2. CZSaw's Hybrid Graph View. The user double clicked on the document "Whitewater, Watergate" to retrieve a set of people nodes (blue nodes) contained in that document.

The Document View (DV) displays data document text and a list of entities contained within all documents (Figure 3-3).

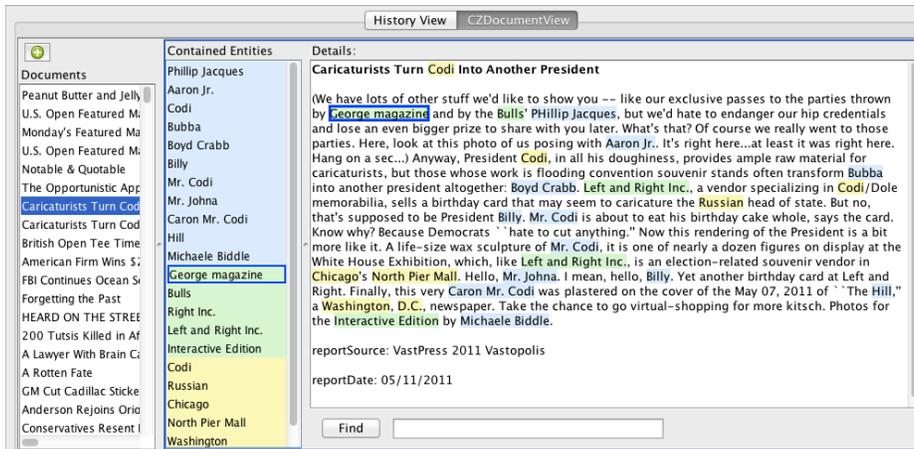


Figure 3-3. CZSaw's Document View – The left panel is a list of CZSaw data documents, the middle panel lists all the entities contained within the selected document, and the right panel displays the text of the document with the extracted entities highlighted.

Aside from the data visualizations, a main focus of *CZSaw* is to capture the analysis process. *CZSaw* captures all of a user's interactions in a script, which can be replayed through *CZSaw*'s scripting engine. This allows the user to review previous analyses and also to share the script of their analyses with others.

Furthermore, the user's interaction with the data builds up dependencies among the results generated by searches and other queries, and this is captured in a dependency model. Such a dependency model has been demonstrated as a way to reuse past analysis processes by re-executing the model on new data (V. Chen, Dunsmuir, Kadivar, & Lee, 2008).

To conclude our *CZSaw* overview, we note the *CZSaw* team was awarded the IEEE VAST 2010 Mini Challenge 1 award for "Outstanding Interaction Model" (V. Chen, Dunsmuir, et al., 2010). It was also awarded, jointly with two other SFU teams using other visual analytics tools, a Grand Challenge award for "Excellent Student Team Analysis" (Dunsmuir et al., 2010).

3.2. Design Approach

Before we begin the discussion of the design of *CZNotes*, we need to define what constitutes a "note". In general, a note can be any artifact created by an analyst to support their sense making, such as a text note, graphical annotation, highlights, etc. For this initial exploration however, we restrict notes to be text documents only.

CZSaw lacked note taking capability to support sense making when it was initially developed. In the initial design of *CZNotes* as a text object within *CZSaw*, we noticed a resemblance to *CZSaw* documents – text with entities extracted. We thought it would be potentially very helpful to analysts to also enable such entity identification and refinement capabilities within *CZNotes* since notes are often written about the entities and the data documents being analyzed. Thus, the simplest method for incorporating *CZNotes* into *CZSaw* was to represent notes in the same way as *CZSaw documents* are represented. As mentioned earlier, *CZSaw documents* are also represented as entity

objects with entity type “report”. Thus *CZNotes* are also represented as entity objects with the entity type “note”.

The implementation of *CZNotes* is simplified with this opportunity of representing *CZNotes* as *CZSaw* documents. In addition, this enables the user to do much more with their notes, such as utilizing the existing *CZSaw* data views to interact with them. We next describe the various properties and functionalities that *CZNotes* inherits from a *CZSaw* document and also the additional interactions we provided in *CZNotes*.

The following are the main design features of *CZNotes*. Most of these features of *CZNotes* are inherited from the properties of a *CZSaw document*. Users can use the same *CZSaw* tools to interact with *CZNotes* as they could with any other documents and entities:

- Visualize and interact with *CZNotes* in existing *CZSaw* Data Views.
- Query *CZNotes* with existing *CZSaw* search capabilities.
- Inherit the Entity Refinement capabilities of *CZSaw documents*.
- Incorporate *CZNote* interactions (i.e., creation and editing a note) into the *CZSaw* Script.
- Integrate *CZNotes* into the dependency model.

Although much can be accomplished using existing functions of *CZSaw* with *CZNotes*, we also needed to add some unique *CZNote* interactions to support note taking tasks:

- Create and Edit a *CZNote*.
- Support for entity recommendations.
- Create a note repository.

By integrating *CZNotes* with *CZSaw* and representing notes as data, several potential benefits accrue to help the user's analysis.

1. Note Reminders in Query Results. By integrating *CZNotes* with the data, we can search notes and the data simultaneously. Combining the query of data and notes and returning notes in the query results along with data results can remind the user of their previously created notes.
2. Support for iterative Analysis. Notes can be a trigger to further analysis. Users can use notes and the relationships among their contained entities to query for other related entities and documents.
3. *CZNotes* can support the creation of new relationships among entities. Entities and documents are related based on bibliographic coupling and co-citation relationships. In *CZSaw*, for example, entities contained in the same document are marked as related. Therefore, an analyst can link two entities together by adding them to a *CZNote*.
4. *CZNotes* can bring new data into *CZSaw*. Analysts can write any type of text content in a *CZNote*. The analyst potentially could refer to information outside of the current data set in the *CZNote* and identify new entities the note. Thus, such new information and new entities could be integrated with the current data set as official *CZSaw* objects.
5. Analysts can use notes in new ways in *CZSaw* Data Views. By enabling users of *CZSaw* to interact with *CZNotes* and the data simultaneously in the existing *CZSaw* data views, analysts can potentially use notes in new ways. We will observe if this is the case in a user study (Chapter 4).

Our goal in representing notes as data is to benefit analysts in their analyses. For example, feeding notes back into the knowledge set (data and synthesized notes) would assist the analyst in the iterative analysis process. We also conjecture that visualizing notes and data in the same data views can benefit the analyst by enabling the visualization of data and notes in the same space to form a more cohesive picture of the facts (data) and synthesized knowledge (*CZNotes*). We are also interested to see if analysts may in fact develop new usage scenarios of *CZNotes* as data. We outline the following possible scenario to illustrate the use of notes.

An analyst initially starts her analysis to uncover a terrorist plot by using the *Semantic Zoom View* to determine the major themes of the data documents (news articles, communication transcripts, and intelligence reports). She notices two prominent people named “Nicolai” and “Boonmee” within a cluster label. She creates a new *CZNote* reminding herself to find more information about these two people. During her analysis, she creates *CZNotes* containing hypotheses regarding the relationship of Nicolai and Boonmee. In one instance, she created a new *CZNote* about Nicolai and Boonmee staying in Nigeria for a few days. The analyst then wraps up for her weekend. At the start of her following workweek a colleague informs her of a virus outbreak and the analyst focuses her attention analyzing the data set for the cause of the outbreak. She investigates the location entity “Nigeria” as a possible source of outbreak by adding the entity “Nigeria” into the *Hybrid Graph View*. Double clicking on the Nigeria entity node triggers a query to find related entities. The results of her query returned Nicolai and also all the previous *CZNotes* that she had written about Nicolai in Nigeria and this helps her to forge the connection between the hypotheses she had written about Nicolai’s terrorist activities and the virus outbreak. The script recording the creation and editing of *CZNotes* was then shared and loaded into another collaborating analyst’s *CZSaw* machine for follow up analysis and review. The collaborating analyst re-executes the script and loads all the *CZNotes* into the *Semantic Zoom View* and runs the clustering tool to determine the theme of the shared *CZNotes* and continues from there.

In the following section, we describe the details of how we integrated *CZNotes* into the existing *CZSaw* system.

3.3. Integrating *CZNotes* into *CZSaw*

3.3.1. *Entities and CZNotes*

When data is imported into *CZSaw*, natural language processing tools are applied to the data set to extract entities (i.e., Named Entity Extraction - NER). The NER process tags text strings within each data document to identify possible people, places, time, etc. What is tagged in this process depends on the type of NER tool used and on the “training set” used for the tool. These entity objects are the core component of *CZSaw*’s data model. An entity has an entity type and a set of attributes. For example, a *CZSaw document* is an entity with entity type “report” and an attribute is the title of the document. Another attribute would be a list of entities contained within the document. Examples of other entity types are people, location, organization, and dates. Other entity types can be pre-defined and used during the entity extraction process – during import of data into *CZSaw*. When a user creates a new *CZNote*, the note is added to this index as an entity with type “note” that has attributes similar to those of “report” entities such as title, text content, and the entities it contains. These entity objects are stored in an index, which can be searched during the runtime of *CZSaw*. In essence, we have integrated notes and the data into the same storage and index. Interactions with the data and notes such as queries will go through this index. This index is built and stored in memory for quicker query processing.

NER, however, cannot in general accurately identify all entities within the text. Therefore, *CZSaw* enables the user to make refinements to the entity extraction process. The user can extract new entities, remove incorrect entities, and merge entities in the documents. These operations can similarly be applied to the text and entities within a *CZNote*. Allowing this capability within a *CZNote* opens up possibilities in terms of how the user can interact with the data set. For example, *CZNotes* could be used as a vehicle to bring external data into the existing data set since the user can write his/her own content in a *CZNote* and then subsequently extract entities from this content.

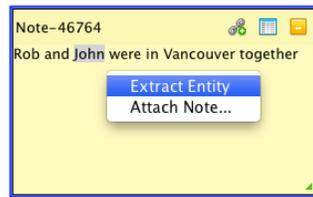


Figure 3-4. A popup menu to extract a new entity from highlighted text within a CZNote.

For example, a user using CZSaw to analyze a set of data may have found relevant information that was not in the original data set; for e.g., about a person named “John” from an Internet search engine. The user can create a note that contains information from the Internet search (i.e., information external to CZSaw) about John and extract John as an entity of type “person” using CZSaw’s entity refinement capabilities. Now John is an “official” entity in the data set. Future queries for all “person” entities would include “John” as a query result. Future queries for “John” can also reveal that there is a note about him.

CZNotes can also reference CZSaw documents and CZNotes as they are also represented as entities. CZNotes can thus be used as a mechanism to build a deep level of interconnectivity among CZSaw documents, other CZNotes, and entities.

Entity Extraction Suggestions

When an analyst creates a new CZNote, the content is initially recorded as plain text. The analyst would need to manually identify any entities within the text content as described in Figure 3-4. To assist the user in identifying entities we developed a feature to suggest entities to reference to the user within the CZNote. This feature scans the content of the CZNote for any text strings that match existing entities in the data set and lists them in a side panel (Figure 3-5). The user can select entities that should be extracted from the CZNote within this side panel. In the future, CZNotes will automatically apply NER tools to extract entities after a note is created.

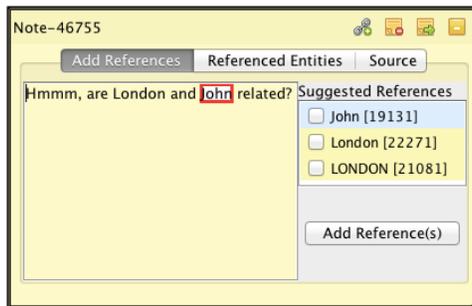


Figure 3-5. “Suggested References” feature is used to scan a note’s text for entities to identify and reference.

Creating Relationships Among Entities using CZNotes

The relationship between documents and entities is defined by containment, i.e., using Bibliographic coupling and co-citation. A document is related to an entity if the document contains the entity. Two entities are related to one another if they are contained in the same document. This model of relationships was originally inspired by the tool *Jigsaw* (Gorg et al., 2007).

A *CZSaw* document Doc1 containing entities C, D and E will have a containment relationship between Doc1 and the entities. The *CZSaw Hybrid Graph View* visualizes the documents and entities as nodes and the bibliographic coupling, co-citation and containment relationships as links. A *CZNote* plays a similar role as a *CZSaw* document in this model of relationships between documents and entities.

Although the initial set of relationships among the entities is important, analysts also need the capability to add their own relationships among documents and entities based on a combination of their own knowledge and external information. Analysts need a way to explicitly create such relationships in order to allow analysts to continue progressive work with them. The creation of bibliographic coupling and co-citations relationships is possible through the use of *CZNotes*.

For example, an analyst writes in a *CZNote* that “John” and “Rob” were at the same location at the same time after reading two separate documents. The initial data set has no document that describes such a connection between “John” and “Rob”. The analyst then creates a *CZNote* stating that John and Rob were both in Vancouver at the

same time and extracts the people and location as entities (Figure 3-6). Now future queries for people related to John would reveal Rob as a connection, which was not present in the initial data set. Similarly, a query for people associated with Vancouver would return John and Rob.

In this way, we are using bibliographic coupling within notes to *create new relationships*. This synthesized knowledge of the note is added to the integrated pool of knowledge, i.e. the data and notes, and can support future analysis.

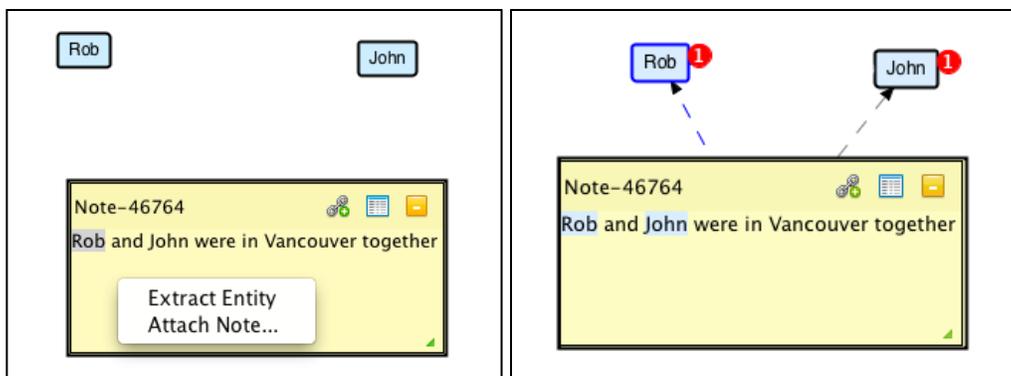


Figure 3-6. The images above show the creation of a relationship between Rob and John through a user created CZNote by identifying the two entities in the same note.

3.3.2. CZNotes and the CZSaw Script

CZSaw captures meaningful user actions in a script (Figure 3-7). The scripting engine also records the action of creating and editing a CZNote. To accomplish this, new scripting commands were added to the CZSaw scripting language to enable the recording of note interactions. The note creation command requires parameters: *Note Title*, *Note Content*, *List of References*, and *Source information*. The *List of References* refers to the entities that the note contains. The *List of References* refer to the entities contained within the CZNote. The *source information* refers to how a note was created. The script allows saving CZNotes over multiple sessions. When a script is opened and executed, the CZNote commands will be re-executed, re-creating the notes.



Figure 3-7. A screenshot of the script interface containing a list of user actions that were previously recorded.

3.3.3. CZNotes and the Dependency Model

CZSaw also creates a dependency graph of the generated data elements as the user interacts with the system. Data elements include the variables storing sets of data entities and data views. Variables containing sets of data entities are created when a query is made, i.e., any query to the data will be stored in a named variable. Any changes to the data will cause an update propagation at that point of the model; consequences of the change are propagated down the dependency graph. By changing the content of a node in the dependency graph, the dependent child nodes will be re-evaluated using the new values in the parent node. This essentially allows the analyst to re-use their analysis process on different data. This mechanism was explicitly used in the 2008 VAST challenge to solve the mini challenge 3 (V. Chen et al., 2008).

The root of the dependency graph is the set of all data in the system. Creating a new *CZNote* adds a note entity to the system data set. Thus, when a new *CZNote* is created or edited, the root node is triggered to update (top blue node in Figure 3-8), resulting in propagation throughout the remainder of the dependency graph. Therefore,

through this dependency model and propagation system, new *CZNotes* and *CZNote* edits will be reflected in the different views of *CZSaw*.

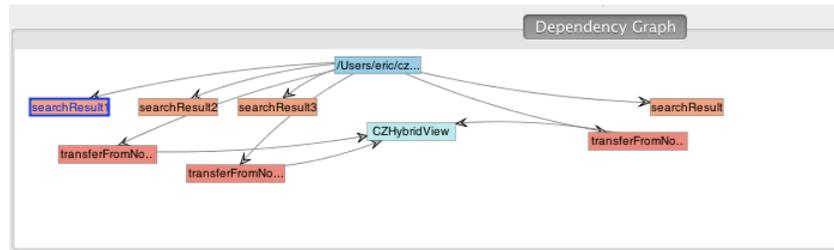


Figure 3-8. A simple dependency graph generated after multiple searches and displaying the search result in a “CZHybridView”. If the user changes the content of the “searchResult3”, the change in content will trigger a propagation to update the contents of “CZHybridView”.

Suppose for example, the user generates a query (we’ll name this query “queryNicolai”) within the *Hybrid Graph View* to get a set of data related to a person named “Nicolai”. This returns a set of documents and notes, (we’ll name this result “resultNicolai”) linked to the person node “Nicolai”. The analyst later creates a *CZNote* about “Nicolai” and identifies him in the note as an entity. This addition of the *CZNote* causes an update to the root node and propagation through the dependency graph, which subsequently re-executes the query to get related data about Nicolai in the *Hybrid Graph View*.

Eventually, “queryNicolai” is triggered to execute again to update “resultNicolai”. In this subsequent query, the *CZNote* about Nicolai that was just created will be visualized in the *Hybrid Graph View* as part of “resultNicolai”.

With this propagation mechanism, new *CZNotes* will be shown in the appropriate contexts, immediately after creation, like in the above example. However, there is no visual indicator if a new *CZNote* has been added to a result set, which is a future work item.

3.4. Note Recommendations and Progressive Analysis

Many sense making and collaborative analysis tools offer ways to reveal relevant notes based on the analysis context as described in chapter 2. Some tools use algorithms to calculate the similarities of note contents to the user's current context such as the current document the analyst is reading (Cheng & Gotz, 2009; Gotz, 2007; Nobarany et al., 2012). Some tools allow users to attach comments, or "notes", to specific states of visualizations (Heer et al., 2007; Willett et al., 2011). If the user returns to this state, the associated comments will be displayed on screen. The goal is to remind the user of their past findings and to continually build up a story of the data and increase the chances of serendipitous findings during analysis.

We hypothesize that with the integration of *CZNotes* and the data, we can achieve a degree of "reminding" similar to the above tools. For example, in the scenario posed above in the "Creating Relationships Among Entities using *CZNotes*" section, the analyst found new information that John and Rob were in the same place at the same time and created a *CZNote* about this. This note creates a link between John and Rob by a co-citation relationship. When the analyst queries for data related to John in the *Hybrid Graph View*, the *CZNote* linking John and Rob in Vancouver is returned. Thus, this novel approach reminds the analyst about their previous findings without adding additional functions to the system. Furthermore, the dependency model described earlier will also update previous user query results to include any new *CZNotes* that match the query.

Our method of including previously created *CZNotes* simply by integrating *CZNotes* as data documents enables this "note recommendation" capability in an unobtrusive and natural manner without dedicated side panels or notifications that could distract the analyst.

3.5. Collaboration - Sharing CZNotes by Sharing Script Files

The script file can be shared among collaborating analysts. This was in fact a crucial step in solving the VAST 2010 Mini Challenge 1; see (Dunsmuir et al., 2010). Since *CZNote* commands are recorded in the script, sharing the script results in sharing *CZNotes*. This script is stored as a simple text file therefore can be edited outside of *CZSaw*.

Creating and editing *CZNotes* adds a new *CZNote* entity to the data model and edits an existing *CZNote*, respectively. Thus commands to create and edit *CZNotes* are considered data manipulation commands (commands that change the data model). *CZSaw* has a functionality to export only the data manipulation commands. In other words, only the *CZNote* creation and edit commands are kept in the script. A benefit of this approach is that a collaborating analyst who wants to utilize only a set of *CZNotes* from another analyst's script does not need to execute all other steps of an analysis. This allows the collaborating analyst to start a new *CZSaw* session or continue or continue an existing session with the knowledge set (*CZNotes*) of another analyst.

3.6. Visualizing and Interacting With CZNotes

CZSaw documents can be visualized in all data views of *CZSaw*: the Document View (*DV*), the *Hybrid Graph View (HGV)*, and the *Semantic Zoom View (SZV)* because they are represented in the same way as *CZSaw* documents. This offers great flexibility and power in interacting with and visualizing *CZNotes* for sense making. Each data view also has an option of adding all the entities of user-selected entity types. Therefore, the user can choose to add all note-type entities into a data view. This lets the user create dedicated "CZNote Views", if desired, by using any of the available data views.

Initially, we considered building a *CZNotes*-specific view for spatially organizing one's notes, similar to Sandbox (Wright et al., 2006). However, we realized that this

spatial organization can be supported using the existing *HGV* and thus a *CZNotes*-specific view was unnecessary. We did however create a *CZNotes*-specific side panel, a “Notes Repository”, attached to the *Hybrid Graph Views* and the *Semantic Zoom Views*, described later.

In the next three sections, we describe how an analyst interacts with and visualizes *CZNotes* in the *CZSaw*’s data views.

The *Document View (DV)* offers a larger viewing area for reading *CZSaw* documents (Figure 3-9). The view is split into three panels. On the left is a listing of all the documents as well as all *CZNotes* added to the view. The second panel lists all the entities that are referenced in (i.e., contained in) a selected document (or *CZNote*) in the left panel. The third, largest, panel displays the document’s text. The user can add all existing *CZNotes* to this view using the “Add All”, or green plus, button in the top left corner). With this add all button, the user can decide to dedicate a view to handling notes.

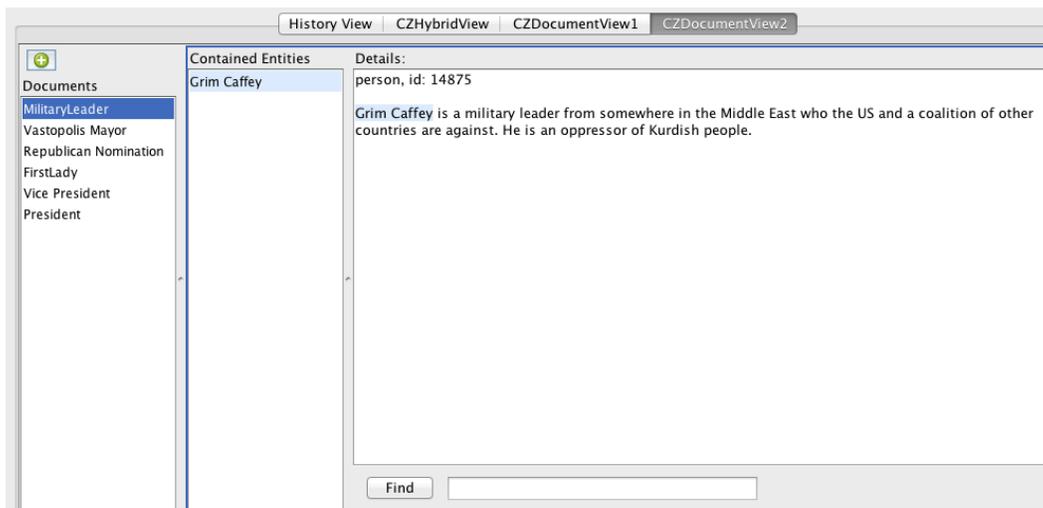


Figure 3-9. The Document View is displaying a CZNote written by the user. All CZNotes can be added to the Document View at once using the Green plus button above the first panel on the left.

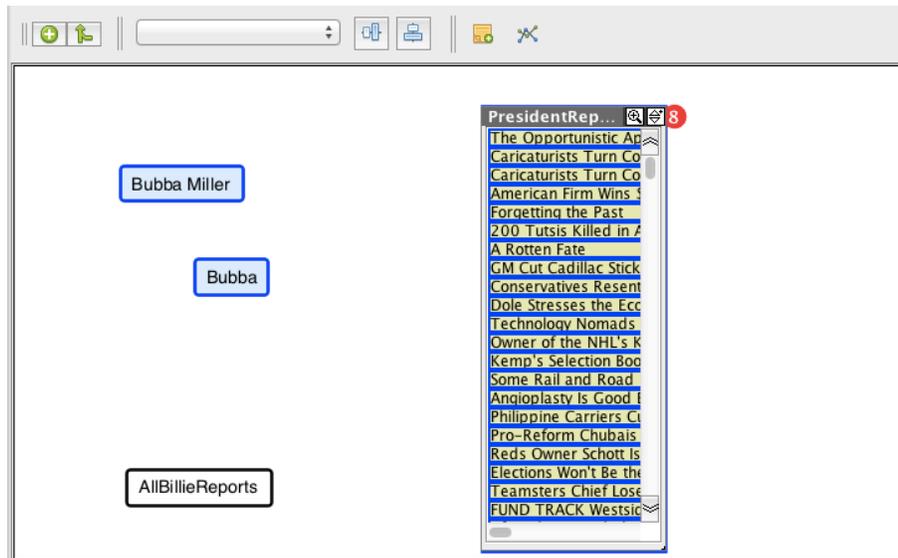


Figure 3-10. This figure shows how nodes can be represented in various forms. For example, the nodes can represent a single person entity node (blue), an aggregate node containing a set of Reports (white node), or as a list node containing a set of entities.

The *Hybrid Graph View (HGV)* enables data entities to be visualized in a node link graph style. Entities can be spatially organized. The *HGV* was given its name because of the data view’s capability to display entities or sets of entities as a single node, in aggregate nodes or as a list node (Figure 3-10). With the integration of *CZNotes*, notes can also be displayed and organized in the same manner as the data within the *HGV*. Many VA tools allow users to create and organize “notes” with a diagramming approach such as *Sandbox* (Wright et al., 2006) and *ARUVI* (Shrinivasan & Van Wijk, 2008) and concept mapping tools. In such tools, the user can quickly create a note and start typing. Our design of *CZNotes* supports quick and easy *CZNote* creation. To do so, we designed an additional node representation specifically for representing *CZNotes* in the *HGV* (Figure 3-11).

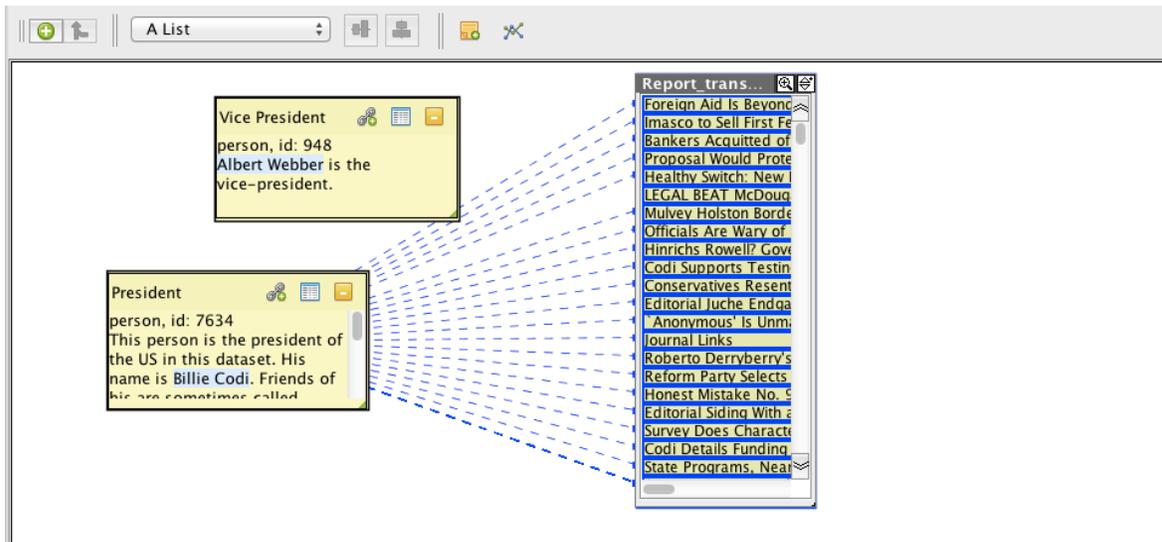


Figure 3-11. The yellow nodes represent CZNotes containing text content with identified entities e.g. “Billie Codi” and “Albert Webber”.

The CZNote node representation in the HGV allows the user to edit the note’s title and content, identify entities, and create connections with other CZNotes and entities. Additional details can be found within the CZNote, such as a list of all the entities it contains, source information i.e. author of the note and where the note was created and how it was created. In the HGV, any node that is referenced by 1 or more notes will have a red bubble appear at the top right of the node and a number indicating the number of attached notes. This note count indicator can prove useful as an indication the node (data) is of interest since a note is referring to it.

A reference to an entity, from a CZNote, does not necessarily require the entity name to exist in the text of the note. Such references can be added to a CZNote by dragging the “Add Link” icon from the CZNote to an entity node or an aggregate node and the CZNote would then reference the chosen entities (Figure 3-12).



Figure 3-12. Creating a link from a note to another entity node by dragging the “Add Link” (icon with a chain and green plus) icon from the source CZNote to the destination node.

Furthermore, the user can interact with the CZNotes as they can with other entity types. For example, a user can double click on a CZNote to retrieve entities and documents related to the note (Figure 3-13). This integration thus allows the user to include their previously synthesized knowledge (CZNotes) in their current analysis.

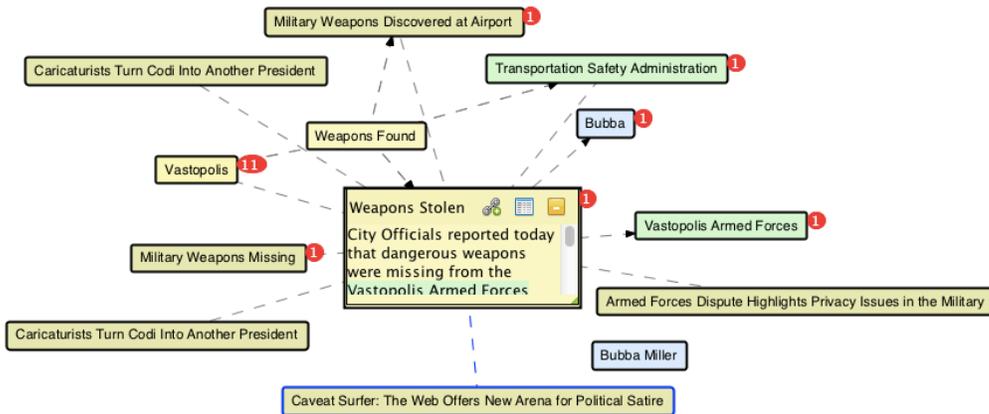


Figure 3-13. Double clicking on the CZNote (from Figure 3-12) retrieves all entities related to the CZNote.

The *Hybrid Graph View* allows the user to freely layout their CZNotes and the data together and to draw connections between the synthesized knowledge (notes) and the “existing knowledge” (the data set). This synergy can potentially provide “aha” moments during an analyst’s visual interaction with the data and the CZNotes.

The *Sematic Zoom View* (SZV) is primarily used to visualize CZSaw document sets. SZV enables spatial placement of documents, clustering of the documents based

on the entities they contain (Figure 3-14), and also semantically zooming into the documents (Dunsmuir et al., 2012).

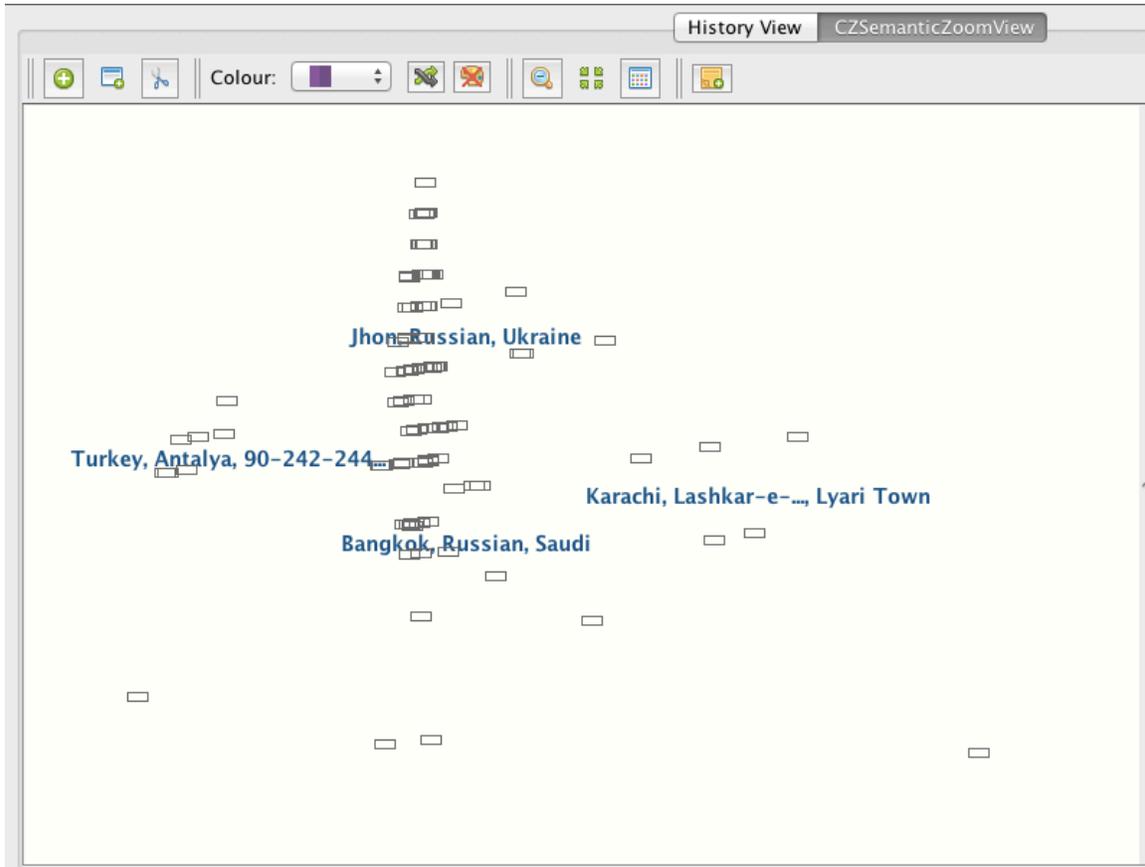


Figure 3-14. The Semantic Zoom View can visualize CZSaw Documents and CZNotes (represented by the grey rectangles). These notes and documents are clustered based on the entities they contain.

CZNotes can be semantically zoomed into just as documents are. The first Zoom level will show the entities referenced within the note as clickable icons. Selecting one of these entities enables brushing and linking, i.e. other documents or notes that contain the selected entity will also be highlighted within SZV. The last zoom level will show the entire CZNote's content, as you would see it in the DV, or in the CZNote node in the HGV. Documents and CZNotes referenced by CZNotes will show a red bubble with the notes attached count, similar to the HGV. The user can utilize SZV's clustering capability as well its zooming capabilities to interact with CZNotes seamlessly with the rest of the

data set. Due to limited resources and time constraints in our implementation, *CZNotes* are not editable in the *SZV* (this capability will be added in future implementations).

A Notes Repository

Although we did not create a *CZNote*-specific view for working with notes, we did develop a notes *repository*, accessible within the *HGV* and the *SZV* as a side panel (Figure 3-15). The notes repository was not added to the *Document View* because it was originally meant for read-only tasks. This panel was designed to provide convenient and quick access to the set of existing *CZNotes* to enable the user to find and add existing *CZNotes* to the *CZSaw* data views.

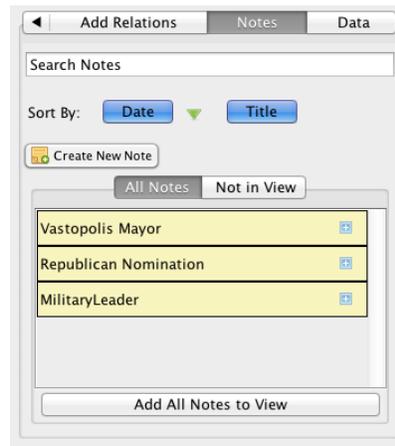


Figure 3-15. The image is of a notes repository containing a list of *CZNotes* that is searchable and sortable.

The note repository provides a list of all notes created by the user. It offers the user the ability to sort, search and filter the list as well as allow the user to drag notes into the data view or add all the notes into the associated data view at once. The list of *CZNotes* is scrollable and each note can be expanded to show its content and enables editing of the note.

To illustrate, consider the following scenario: A user creates a set of notes during one analysis session. He restarts the same project and would like to review a specific set of notes creating a *SZV* and switches to the note repository panel. He filters the list

of *CZNotes* to a set of notes he wants using the search bar. He then drags the filtered *CZNotes* from the note repository into the *SZV* (or uses the “Add All Notes to View” button to add all filtered notes to the view at once).

In hindsight, a better approach would have been to improve upon the existing *CZSaw* search panel and add filtering, sorting, drag and drop capabilities to all entity types rather than build a notes specific side panel. This would provide a simpler and user-friendlier experience than potentially confusing the user with 2 different search panels. We will further describe potential improvements in the Future Work section. There was consideration to remove the notes repository from *CZSaw* at one point; however, we decided to keep it in the system to observe how users would use it.

3.7. Interacting with *CZNotes*

A user can create, edit and delete *CZNotes*. They can also extract and refine the entities within a *CZNote*. We discuss the details of such interactions in the follow sections.

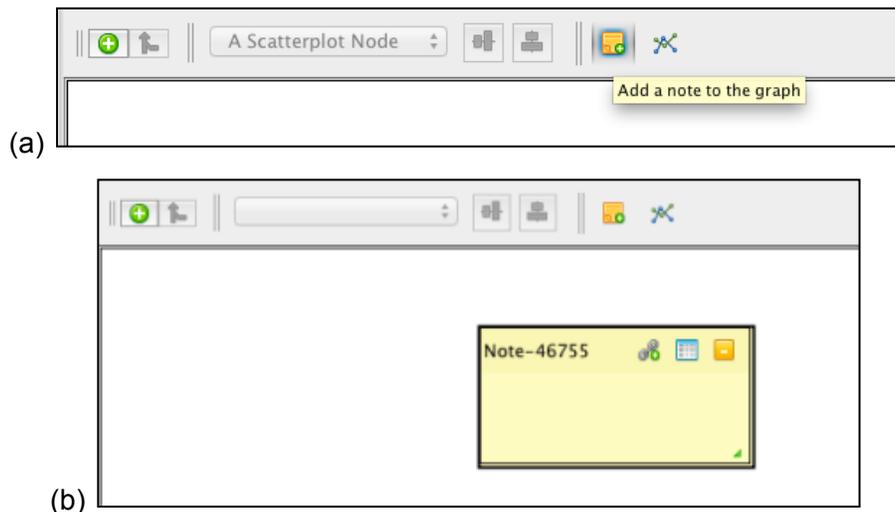


Figure 3-16. (a) A blank note can be quick created in the Hybrid Graph View by toggling the “Create Note” icon into the add note mode (b) and then clicking at a point in the Hybrid Graph View to create a *CZNote* at that location.

The user can create blank notes in the *Hybrid Graph View* by clicking on white space when in “Create Note” mode. The new note will appear where the user clicks (Figure 3-16). Clicking the add note icon in the notes repository also generates a create *CZNote* dialog.

Analysts often want to add information about a specific entity. To support this, *CZNotes* allows users to easily “attach” a note to an entity, i.e. create a new note that mentions (references) one or more entities. *CZNotes* can be attached to entities and documents in all of *CZSaw*’s data views. In the *Hybrid Graph View* and *Semantic Zoom View* can attach notes to select entity/document nodes.

Additionally, notes can be attached to a selected portion of text in a *Document View* by highlighting and right clicking the text and choosing attach note. A “Create Note” dialog box lets the user to fill in the note’s content. In addition *CZNotes* will automatically capture the source information including who created the note and what text was referenced from which *CZSaw* document. The selected text is automatically pasted into the note’s content. The created *CZNote* will add the referred *CZSaw* document to its list of contained entities. This creates a link between the new note and the document.

For example, the analyst finds a passage “explosion that knocked the 747 out of the sky” and wants to make a note about this incident for later reference. The analyst can highlight this passage and attach a *CZNote* to the text (and document). A dialog window for creating a new note will appear with the selected text copied into the note’s content text box (Figure 3-17). The analyst can then make further edits to the *CZNote*.

The “Note Source” is a property of the note that records information regarding the note creation context. For example, the original text that the note is being attached to and the ID and name of the report are recorded as part of the “Note Source”. This source information is currently stored as plain text in the script. In the future, we could use this information to support collaboration and query the notes.

The user can also add the new CZNote to a data view when it is created. Whether or not the note is displayed in a data view, the new CZNote will be added to the CZSaw data index when it is created.

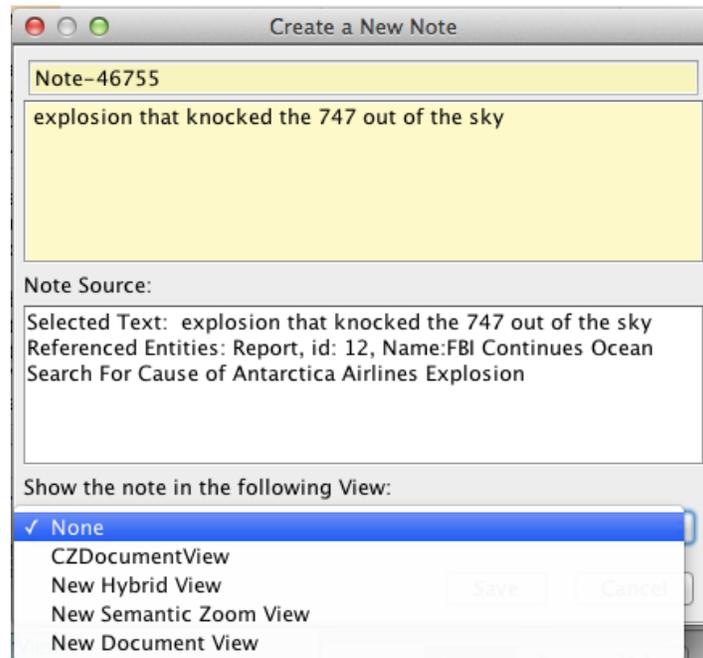


Figure 3-17. A new CZNote is being attached to the Report with ID = 12 with the text selected as “explosion that knocked the 747 out of the sky”. The user can also choose to display the CZNote immediately in an existing data view, a new data view or none.

Similarly, CZNotes can be attached to selected nodes in the *Hybrid Graph View*, whether the nodes are CZSaw documents, entities or CZNotes. For example, an analyst may have searched for entities of type “person” named Nicolai and displayed the results in a *Hybrid Graph View*. This results in two Nicolai’s in the graph. The user may hypothesize that these 2 are the same person and creates a CZNote referencing these two entities for later review (Figure 3-18). In this example, the source information similar to the above *Document View* example is also captured within the new note’s attributes.

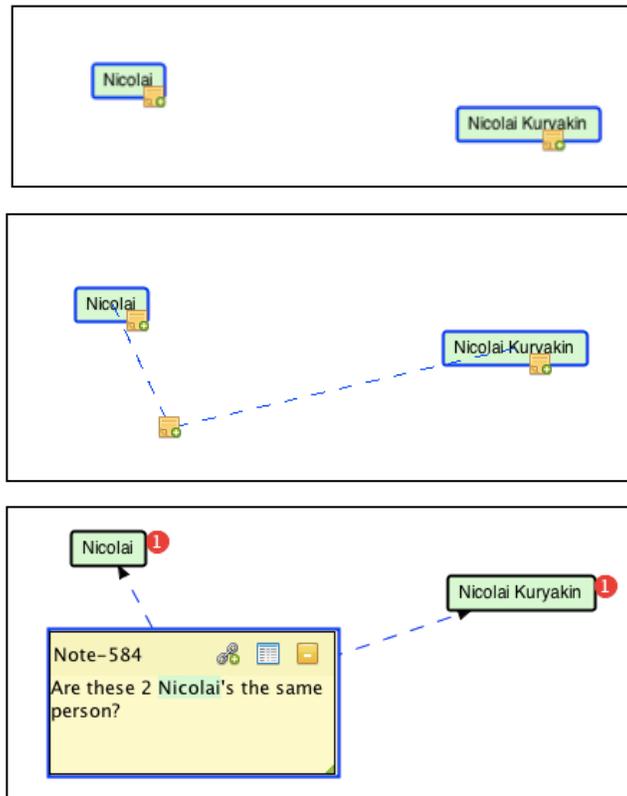


Figure 3-18. A CZNote is being attached to 3 selected entities by dragging the add note icon from one of the selected nodes.

Finally, in the *Semantic Zoom View (SZV)*, new CZNotes can be attached to multiple CZSaw documents and CZNotes in a similar way as in the *Hybrid Graph View*. The SZV uses an “Add Note to all selected “documents” and triggers a “Create Note Dialog” box for creating a new CZNote which references the selected CZSaw documents and CZNotes as references.

For example, the user runs the clustering on a set of CZSaw documents in *Semantic Zoom View* and sees a cluster of documents with an interesting theme titled “low altitude bird strike”. The user selects all the “bird strike” related documents and click son the “Attach Note” button at the top of the view and attach a new CZNote to comment about these documents.

The user can edit CZNotes in the *Hybrid Graph View* when the CZNote node is in “note” mode. (The user can change the node representation by using the drop down at

the top of the *Hybrid Graph View* toolbar). When in note mode, the user can edit the text and the title of the note. Furthermore, the user can highlight and extract entities in this mode. *CZNotes* can also be edited in the Note Repository the same way. The note is automatically saved after a few seconds of inactivity.

Notes cannot be edited in the *Document View* or the *Semantic Zoom View*. Originally, we tried to minimize the alterations of the data views when adding *CZNotes*. Therefore, the *Document View* and the *Semantic Zoom View* was limited to only be used for browsing the data.

CZSaw supports drag and drop capabilities, which *CZNotes* takes advantage of. *CZNotes* can be dragged and added to other view tabs. For example, *CZNotes* can be selected from the *Document View* list and dragged into a *Hybrid Graph View* or *Semantic Zoom View* for further analysis. *CZNotes* can also be dragged from the Notes Repositories into the data views.

3.8. Implementation Overview & Software Architecture

CZSaw consists of Data Visualizations, a History component and a Dependency Graph Model, which all connect to an underlying Scripting Engine. *CZNotes* was integrated into *CZSaw* by representing the *CZNotes* in the same way as *CZSaw* documents. This enables the user to interact with the notes using all the above features of *CZSaw* as one would with *CZSaw* documents. Figure 3-19 *CZNotes* is situated within the *CZSaw* architecture.

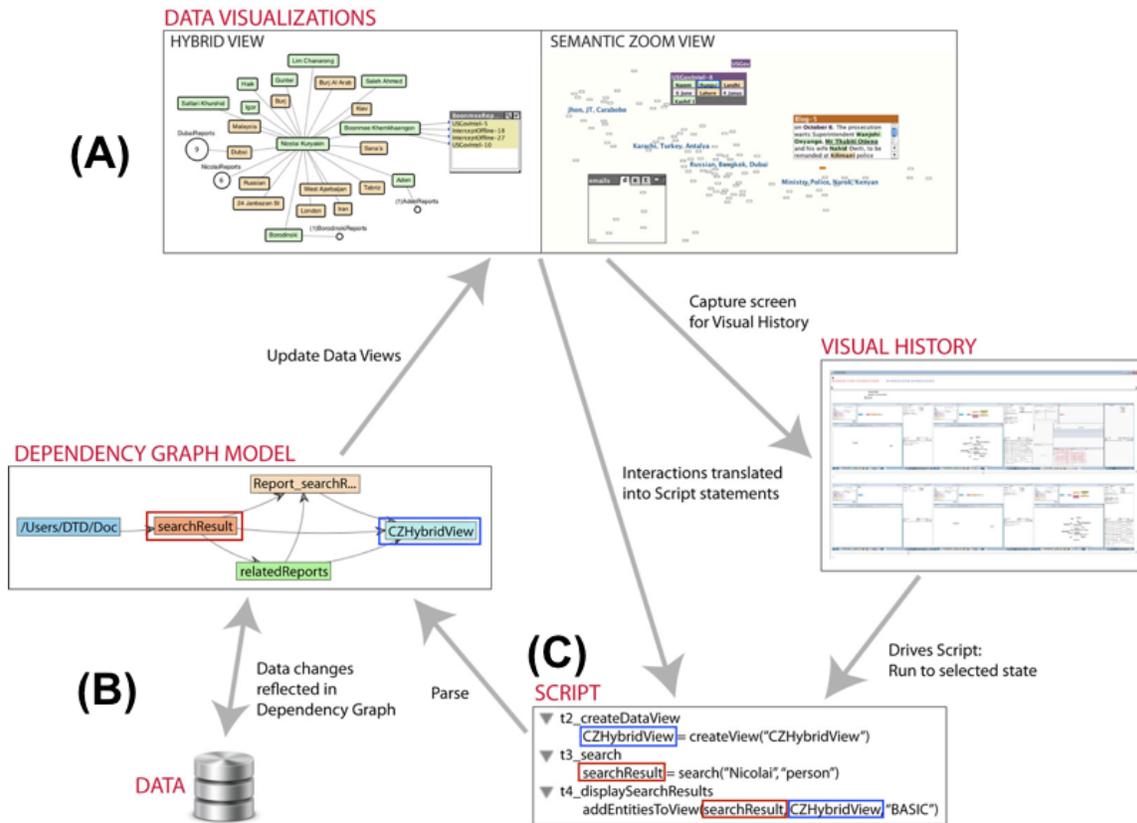


Figure 3-19. The architecture of CZSaw and CZNotes. By representing CZNotes in the same way as CZSaw documents, CZNotes is integrated with all the major components of CZSaw. (A) CZNotes can now be visualized in the existing Data Visualizations (B) CZNotes are represented and indexed similarly as the data and thus incorporated into the Dependency Graph Model (C) CZNote interactions are also recorded in the CZSaw Script

Technology Used

CZSaw was implemented using Java and a number of existing open source libraries. The final application is an executable jar file that can be run on all OS environments with the Java Runtime Environment installed. The *Hybrid Graph View* was implemented using the JUNG graphing library. The SZV uses the library ZVTM for its zooming capabilities. The scripting engine/parser was developed using ANTLR. The entire project is maintained using the Redmine project management tool while the source code is maintained in a GIT repository.

Limitations of Our Design

The user interactions with *CZNotes* are not completely consistent in the current design. We would like to improve the user interaction for using *CZNotes* in the future.

In reviewing the editing capabilities of *CZNotes*, we believe that *CZNotes* should be editable in all data views to be consistent and more useful. However, at the time of this thesis, *CZNotes* are not editable in the *Document View* or the *Semantic Zoom View* (except within the notes repository in a *SZV*).

3.9. Summary of Design

We wanted to integrate *CZNotes* into *CZSaw* in such a way that notes can be part of the data set, allowing the user to query, visualize and interact with *CZNotes* just like they would interact with the *CZSaw* documents. We have incorporated *CZNotes* into the 3 *CZSaw* data views and the user can perform the same actions on the *CZNotes* as *CZSaw* documents. In addition we added note-editing capabilities in the *Hybrid Graph View* and in the new Notes Repository side panels. We have also added the capability to create new *CZNotes* to entities in all 3 data views.

In summary, *CZNotes* was sufficiently integrated into the existing *CZSaw* system to run a user study to observe the usage of *CZNotes* in solving the VAST 2011 Mini Challenge 3 ("IEEE VAST Challenge 2011," 2011). The next Chapter describes the design of the user study and the methods used to carry it out.

4. User Study

In this chapter we describe how we designed and carried out a preliminary user study in order to observe how participants would use *CZNotes* (i.e., notes represented as data) within a VA tool, *CZSaw*, to help solve an analysis task. In this study we investigate how representing *CZNotes* as *CZSaw* data documents can affect the users in their analytical processes. The questions we were interested in answering were:

- Would returning *CZNotes* in query results remind analysts of previously created notes?
- Would *CZNotes* be used to ‘continue’ an ongoing analysis? For example, an analyst might first review *CZNotes* in follow-up analysis sessions or continue their analysis along a certain path when seeing previous *CZNotes* in subsequent query results.
- Would *CZNotes* be used as a means to create relationships among the data through co-citation?
- Would *CZNotes* be used as a way to introduce new data into the *CZSaw* system?

4.1. Study Design

In the real world, analysts work on problems over long periods of time. When evaluating a new VA tool, it is difficult to measure the tool’s effectiveness in a laboratory setting with a limited time. Analysts require time to learn a new tool. Also, an interesting task needs to be provided for the analysts to be engaged. There is no single methodology that can be used to measure how effective a Visual Analytic tool is

because of the complexity of the tool and the fact that the problem being solved may be open ended with no final “correct” answer.

(Redish, 2007) suggests that “[n]o single methodology or measure is going to work for usability testing of all these systems for domain experts doing open-ended problem solving”. Redish et al. suggest the possibility of conducting usability studies outside of the laboratory and “unattended data capture for portions of a long-term evaluation, used along with observations and interviews”.

This approach is also supported by (Shneiderman & Plaisant, 2006) who propose a mixed methods strategy, based on “lessons from ethnographic observations in HCI design”, for evaluating such complex systems, specifically, information visualization tools (Shneiderman & Plaisant, 2006). They suggest a multi-dimensional approach is required with case studies over long periods of time, along with surveys, interviews, ethnographic observation of participants and automated user activity logging.

Based on these suggestions, we employed multiple methodologies, both quantitatively and qualitatively, in a longitudinal study to observe the use of *CZNotes* within *CZSaw*. These multiple means of data collection provided us with a more comprehensive understanding of the use of *CZNotes* and *CZSaw*.

The participants’ task in our user study is to solve the IEEE VAST 2011 Mini-Challenge 3 using the data and task descriptions provided in the challenge (“IEEE VAST Challenge 2011,” 2011). The study consisted of a single team collaboration among 4 participants over the course of 3 weeks to solve the Mini Challenge. The data set consists of 4000 fictional news articles created as a combination of machine generated documents and manually edited content. Participants were provided a written description of the task in which they were asked to use *CZSaw* and *CZNotes* to analyze the news articles and identify and report on any threats to a fictional place called “Vastopolis” based on their analysis (Appendix B). The articles publishing dates ranged over a period of 3 months. We encouraged participants to use *CZSaw* and *CZNotes*, but did not disallow the use of other analysis tools if they were felt to be necessary. Participants did initiate the use of a *wiki* as a central collaboration space as well as email and instant

messaging to communicate. Participants were strongly encouraged, but not required, to first create a *CZNote* with the information that was going to be shared on the *wiki*. The participants were allowed to work on the challenge at any time and place and with any other participants.

The user-study began with an online pre-study questionnaire (Appendix C) followed by a briefing session where the study details were described and were given an introduction to the features of *CZNotes*. A video tutorial and a PDF tutorial about *CZNotes* were available for participants to use. The participants watched the video tutorial during the first briefing session. However, we did not confirm if the PDF tutorial was used or if the video tutorial was watched again. The participants had 3 weeks to collaborate to solve Mini Challenge 3. During these 3 weeks, the participants and the investigator met 2 to 3 times a week in discussion group sessions to discuss the analysis progress and obtain their feedback on *CZNotes*. Participants answered a post-study questionnaire (Appendix D) at the end of the 3-week user-study.

The participants were all visual analytics researchers: one Masters student, two PhD students, and one Post Doctoral Fellow. All were also members of the *CZSaw* research group. We chose this group of participants because it was not possible to recruit professional analysts who had time for the study and sufficient familiarity with *CZSaw* and *CZNotes* to use them effectively.

(Redish, 2007) suggests domain experts must be involved in "the planning, design, and development of ... systems". She suggests that false negative results may occur without the participation of domain experts. While we acknowledge the importance of involving domain experts in usability testing as Redish suggests, we simply did not have access to such experts. We note this is the case for many if not most VA systems developed in academic environments. Rather, our participants were knowledgeable about the higher-level ideas and design of *CZSaw* because they were designers and developers of the tool itself, but not necessarily of all the features of *CZNotes*. The participants' involvement in VA research meant they had an understanding of the analysis process and had used other visual analytics tools. One participant had previously participated in the IEEE VAST 2010 Challenge and was a member of a team

that won an award for Mini Challenge 1 and an award for the Grand Challenge (V. Chen, Dunsmuir, et al., 2010; Dunsmuir et al., 2010).

We used the following methods to observe and collect data for this user study. Further details are given below.

1. Pre- and post-study questionnaires.
2. A system log of participant's interactions with *CZSaw* and *CZNotes*.
3. Discussion group sessions during the study.
4. Participants were asked to keep a journal for each analysis session.
5. A final report of their analysis written by the participants.

Pre-Study Questionnaire

Participants completed a pre-study questionnaire to gather information on their background experience and their understanding of the analysis process (Appendix C).

System log of participant's interactions with CZSaw and CZNotes

As described in previous chapters, *CZSaw* captures a history of semantically meaningful actions and transactions that a user performs and stores this history in a script file (a .txt file within a *CZSaw* project folder) that represents the current state of the analysis. Semantically meaningful actions include those that act upon the data model and those that modify the data views.

The script files users generated from their use of *CZSaw* were collected for analysis. These script files enabled the users and the researcher to reload past analysis sessions. This script file was useful to allow the users to reload past analysis sessions and continue their analysis.

In addition, the analysis of the users' use of *CZNotes* and *CZSaw* required information not captured in the current script. We wanted to capturing how *CZNotes* were used during analysis, for example, "How were the *CZNotes* returned in search results used?", "How were *CZNotes* visualized?" and "Were new relationships between entities created through a *CZNote*?" Since the current script only captures *CZNote* creation and *CZNote* edit actions, we instrumented *CZSaw* to output such *CZNote* interactions into a separate *CZNote* log file. A table of all such tags are outlined in Appendix A.

Whereas the *CZSaw* script only represents the current state of the analysis process, the *CZNote* log file recorded the chronological history of all interactions performed by the user, such as the opening and closing of a project or replaying the script. Each of these actions was also time stamped in the *CZNote* log file. This log file could not be re-executed though in the future, we may consider adding this capability.

The *CZNote* log files were analyzed to extract frequency counts for the different ways *CZNotes* were used, e.g. the number of new entity relationships created by using a *CZNote* or how often *CZNotes* were returned as part of search results.

Discussion Sessions

Participants met with the investigator seven times during the 3-week period, twice per week for the first two weeks and three times in the last week, in discussion group sessions. Generally, the first half of each session was made available for the participants to collaborate on the analysis task itself and share their findings, describe their processes and plan next steps, and provide their feedback on their usage of *CZNotes*. In the latter half of each session, the investigator asked more focused questions to gather feedback about specific *CZNotes* features, usability issues, etc. These focus sessions were video and audio recorded. (See Appendix F for participation consent form).

Journaling

Participants were asked to keep a journal of their activities throughout the study regarding their activities each time they worked on solving the challenge. The journal was guided by a set of questions asking the participants what the goals of their analysis session were and their feedback on *CZNotes* usage. Participants were allowed to submit their journal entries by whatever means they were most comfortable with. All participants submitted electronic journals, either through a web survey the investigator prepared or as a text file that the participant maintained. The questions participants were requested to address in their journal entries were (also found in Appendix B):

1. Enter your participant ID
2. Please enter the date of this analysis session
3. What did you do in this analysis session? What was your goal? What was accomplished?
4. Collaboration (who, what, when, where, why, how?)
5. How did you use *CZNotes*?
6. Did you perform any entity refinements? Where and what type (e.g. merging entities in a document, extracting a new entity in a note)?
7. Do you have any other comments or feedback about *CZNotes*?

Final Report

Participants were asked to follow the guidelines of the IEEE VAST 2011 Mini-Challenge 3 solution submission guidelines and prepare a report describing their solution. Only 2 of the participants worked on this report because of the limited availability of the other two participants' at that time.

Post-Study Questionnaire

We also followed up with participants regarding their familiarity with other visual analytic tools following the study (Appendix D). Participants also completed a post-study questionnaire to provide feedback about *CZNotes*. The questions asked pertained to the likes and dislikes of specific *CZNotes* features and any additional feedback the participants wanted to provide (Appendix E).

In chapter 5, we present how the quantitative and qualitative data were analyzed.

4.2. Data Analysis

The *CZNote* log files provided a source of quantitative data for this study. The log file captured the actions the users made in *CZSaw* and *CZNotes* during their analyses. We used PERL scripts to parse this log file to extract the frequency of the actions along with the timestamp for each action and outputted as a CSV file for further processing (See Appendix G for a sample PERL script).

To accurately extract the number of times each transaction type was executed, we needed to take into account that the *CZNote* Log captures script re-execution. When the user replays a script, the *CZNote* Log captures the replay of transactions within a start and stop tag. Therefore, to avoid duplicating counts of user actions, we removed all the transactions within the re-execution tags from each participant's log files. Figure 4-1 shows a script being re-executed.



Figure 4-1. A CZSaw script executed to the transaction t3_note: transactions below the line at t3_note have not yet executed.

We then used a PERL script containing a set of regular expressions to extract each transaction name and associated timestamp from the log file, adding an entry to a CSV file such as shown in Table 4-1.

Transaction Name	Log File Name	Timestamp	Participant ID
createNote	log 4.txt	2011-09-02 17:22:53	P4
editNote	log 5.txt	2011-09-02 15:34:48	P4

Table 4-1. Sample of transactions outputted to the CSV file.

Similarly, another PERL script was used to parse CZNote tags from the CZNote log file into a CSV file shown in Table 4-2.

CZNote Action Tag	Count	Log File Name	Participant Id
search panel notes-only results	1	log 11.txt	P1
new-relationship created	2	log 14.txt	P1

Table 4-2. Example rows of data pertaining to the additional CZNote tags outputted to the CSV file.

We then imported the resulting CSV files into Tableau (Tableau Software, 2011) for further analysis. Tableau was used to determine frequency counts of user actions made throughout their analyses. With Tableau, we can quickly determine the actions made by each participant and the distribution over time.

From the 3-week study, participants generated between 3 to 20 log files each. The number of log files does not correspond to the number of sessions a participant had since participant may have generated only one *CZNote* log file if the participant only used one *CZSaw* project for the entire study (One log file is created per *CZSaw* project). The CSV data obtained from the parsing of the logs were loaded into Tableau for further analysis, after first removing all non-*CZNote* transactions. We describe the results of this analysis below.

Discussion group sessions were both video and audio recorded. However, since no additional significant information was found in the video recording (participants did not present their analysis progress with their laptops or show any visuals during the discussion group) only the audio recordings were transcribed, with a separate transcript for each discussion group session.

We used (Bartas Technologies, 2011) to assist in manual transcription of the audio in a structured way. The software enables keyboard shortcuts to identify the speaker, the time, and also the dialogue into an utterance, i.e. a *Speaker, Time, Dialog* row in the coding file. Each focus group session transcription is associated with that session's audio file. The coding tool links the time of a transcription to the place in the audio file. Following the initial transcription, grammar and spell check in Microsoft Word was used to scan the transcribed text documents. Any grammatical errors that the spell check revealed or any confusion on what was said was double-checked by reviewing the original audio file. Each session's transcriptions were exported as a simple text file with a list of all utterances.

We analyzed the discussion group transcriptions and journals using a coding approach by categorizing the participants' feedback into themes (or codes). We began coding of the transcripts by first using an initial set of codes (Table 4-3) based on a set of initial research questions. As we progressed, we added codes as new themes emerged

from the data. After coding the transcripts, we used the resulting code set on the journals. This coding the journals and transcripts was an iterative process.

Code	Research question
REMINDERS	Will analysts be aware or reminded of their previous findings (notes created by themselves or others) during their analysis with <i>CZNotes</i> being returned in query results?
CONTINUED ANALYSIS	Will <i>CZNotes</i> be used as a starting point for continued analysis?
CREATE RELATIONSHIPS	Will <i>CZNotes</i> be used as a means to create relationships among the data through co-citation?
EXTERNAL DATA	Will <i>CZNotes</i> be used as a way to bring external data into the system (<i>CZSaw</i>)?

Table 4-3. The initial set of codes based on the initial set of research questions.

The focus group session transcripts (exported from Transcriva 2) were coded using the Atlas.ti Qualitative Analysis Tool (ATLAS.ti Scientific Software Development GmbH, 2011). The first coding pass involved adding two attributes or “tags” to each <Speaker, Time, Dialog> “row”, i.e., to each utterance. The two attribute tags were *speaker* and *topic*, with the topic being an Atlas.ti code that was created (subjectively) by the investigator, whom categorized each utterance. Each utterance could be assigned 1 or more codes (Table 4-4). These codes were related to *CZNotes*, *CZSaw*, the analysis process, findings in the data, or performance issues and bugs. Coding all the focus group session transcriptions produced a total of 164 codes used for the 951 utterances. All codes related to *CZNotes* were gathered in an Atlas.ti’s network view to group and refine the set of codes - that is, to group similar codes together and merge duplicate codes. This process resulted in the final set of codes shown in Table 4-4. The journal text was subsequently coded similarly using the final set of codes generated from the focus groups.

Code	Description
CATEGORIZE	<i>CZNotes</i> as a Categorizing Tool
COLLABORATION	How participants shared their <i>CZNotes</i> with one another.
CONTENT	The type of content written in <i>CZNotes</i>
CONTINUED ANALYSIS	Using <i>CZNotes</i> to continue one's analysis.
CREATE RELATIONSHIPS	Using <i>CZNotes</i> to create new relationships among the data.
DATA LINK	The linking <i>CZNotes</i> with the data.
ENTITY REFINEMENT	Refining the entities within <i>CZNotes</i> .
EDIT NOTES	How <i>CZNotes</i> were edited.
EXTERNAL DATA	Bringing in external data through <i>CZNotes</i> .
EXTERNAL NOTES	Notes created outside of <i>CZNotes</i> and <i>CZSaw</i> . E.g. on paper or a text editor.
ISSUES	Issues, bugs or feature requests for <i>CZNotes</i>
REMINDERS	How participants were reminded of past findings through the use of <i>CZNotes</i> .
VISUALIZING NOTES	Organizing or Visualizing <i>CZNotes</i>

Table 4-4. The table represents the final set of codes created during our analysis.

In Chapter 5, we describe the results obtained from the study. Due to the large amount of data gathered, we focus only on a subset of the data gathered.

4.3. Limitations

The principal investigator was the sole coder of the qualitative data. Ideally multiple and multiple coders would analyze the data ensure inter-rater reliability. However, we were unable to find additional transcribers and coders due to limited resources in time to transcribe the group sessions multiple times and in finding a coder who understood the context in terms of *CZSaw* and *CZNotes* and the analysis process well enough to accurately transcribe and code the information.

5. Results

In this chapter, we combine the results of our quantitative and qualitative analysis to describe the usage of CZNotes. Due to the large amount of data gathered, we will focus on the results of the following areas:

1. CZNotes support for continued analysis.
2. Using *CZNotes* to link to the *CZSaw* data.
3. Use of *CZNote* as a reminder of past findings.
4. Use of *CZNotes* to create new relationships.
5. The introduction of new data into *CZSaw*.
6. Visualizing *CZNotes* in *CZSaw*'s existing data views.
7. Using *CZNote* as a categorizing tool.

5.1. Questionnaire Results

First we describe the results of the pre-study questionnaire and the VA tool familiarity questionnaire. All participants have an academic background in researching visual analytics, ranging from an MSc student to a Post Doctorate Fellow. Three participants had intermediate experience with existing visual analytic tools such as *Tableau*, *In-Spire* and *Jigsaw*. In terms of experience with *CZSaw*, two participants were highly knowledgeable of how to use *CZSaw*. The other two participants had intermediate to low experience using *CZSaw*.

In the pre-study questionnaire we also asked participants what their past experience was using notes to solve problems. Three of the participants did use notes in

the form of wiki posts, notes in *PDF* reader applications, and also “memos and key words for qualitative analysis of interview transcripts.” The use of the notes were to record a history of what had been done, to organize hypotheses, to collect findings about the data, and to keep track of concepts. These notes were also shared with others. P2 said no notes were used or shared in previous problem solving tasks.

The participant’s responses to the post-study questionnaire are included in the relevant sections in this chapter. Regarding the question “Did you find it beneficial for notes to be represented the same way as the data (i.e. *CZNotes* as another entity type)?” in the post-study questionnaire, 2 participants responded that it was “Somewhat Useful”, 1 participant responded “indifferently” and 1 participant said it was “Somewhat Not Useful”.

5.2. *CZNotes* Usage Statistics

Over the course of the 3-week study, participants created 103 *CZNotes*, edited *CZNotes* 138 times and appended entity references to *CZNotes* 93 times (Table 5-1). Participants P1, P3, and P4 were the primary contributors to these *CZNote* actions. P2 created 1 *CZNote* during the study, perhaps due to P2’s very busy schedule during the time frame of the study.

Action	Count
<i>CZNotes</i> Created	103
<i>CZNotes</i> Edited	138
Entities Added to <i>CZNotes</i>	93

Table 5-1. The total number of *CZNotes* created, and edited by all 4 participants.

It was not possible to track the deletion of a *CZNote* or the removal of an entity reference from a *CZNote* because the same transaction name, “removeEntity”, is used for the deletion of an entity for any entity type, i.e., *Person*, *Location*, *Date*, etc., including

the *Note* entity type. Therefore, we could not distinguish when the `removeEntity` command was used for the deletion of a *CZNote*.

Participant ID	CZNotes Created	CZNotes Edited	Entities Added to CZNotes
P1	59	131	88
P2	1	0	0
P3	15	7	5
P4	28	0	0

Table 5-2. The number of CZNotes created per participant, the number of times CZNotes have been edited, and the number of times references to entities have been added to a CZNote.

In the group discussions participant P4 stated he would like to be able to easily edit an existing *CZNote*. From Table 5-2, we see P4 did not edit his notes at all. This could be due to usability problems with editing and adding references to existing *CZNotes* as described in P4’s qualitative feedback described below.

5.3. CZNotes Support for Continued Analysis

Participants described different *CZNote* usage scenarios depending on whether they were discussing the foraging loop of an analysis or the sensemaking loop as described by (Pirolli & Card, 2005). Participants continued their analysis from previous sessions by reloading the *CZSaw* script containing *CZNotes* from their past session or from another participant’s shared script. We describe how these sequential analysis sessions worked.

5.3.1. CZNotes and the Foraging Loop

In the beginning of the study, participants explored the data without a specific thread to follow, except for the contest’s main goal. Participants used *CZNotes* similar to Pirolli and Card’s description of a shoebox by creating “quick and dirty” *CZNotes* and

adding any interesting data to them as references. P4 and P3 described collecting *CZNotes* on various data facts or interesting findings in *CZNotes* when in “exploratory mode” and “splatter[ed] notes all over the place”, said P3. They created *CZNotes* with a minimal amount of text, essentially using *CZNotes* as a container or a tag to tag the data, e.g., analogous to pasting sticky notes onto a paper document.

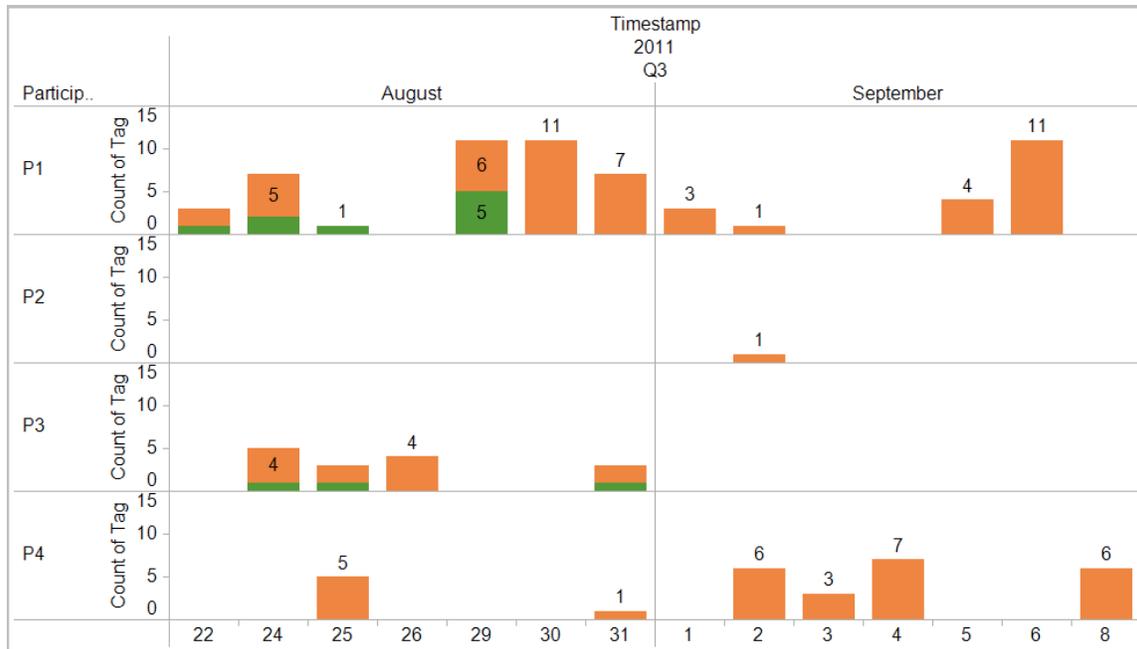


Figure 5-1. Total number of notes created over the course of study. Green bars: Creation of a new blank *CZNote*. Orange bars: A *CZNote* was attached to a *CZSaw* object (a document or to another entity). (See Chapter 3).

Participants created *CZNotes* primarily by attaching new *CZNotes* to documents and entities rather than creating a blank *CZNote* without any references (Figure 5-1). This may suggest that when foraging in the data participants wanted to create *CZNotes* related to a source. It seems that no new blank notes (green bars in Figure 5-1) were created in later analysis stages.

P1 said, “I wanted to add notes to any documents or entities that I found interesting”. P1 also said in one of his journal entries, “I did not do that much editing or organizing of existing notes as I am just trying to uncover lots of facts for now and will

sort them later.” This further supports that *CZNotes* were primarily used as a “shoebox” during the early part of analysis.

Attach <i>CZNote</i> in:	Count
<i>Document View</i>	75
<i>Hybrid Graph View</i>	3
<i>Semantic Zoom View</i>	13

Table 5-3. A count of how many times *CZNotes* were attached to data in each of data views during the course of the study.

The log indicates that *CZNotes* were primarily attached to documents when reading through documents in the *Document View* as shown in Table 5-3.

By replaying one of P4’s scripts we observed the process for attaching *CZNotes*. In Figure 5-2 (a) and (b), P4 opened a *Document View* and added all the new article reports to it. He then scanned the reports and attached a *CZNote* to any interesting entities. He did not add these newly created *CZNotes* to a data view suggesting he used the notes primarily for record keeping purposes.

Creating too many *CZNotes* at this stage of analysis apparently was not a major concern. P4 stated he was “not too worried about useless notes because it’s pretty easy to identify them and just not pay attention to them.” P4 would capture anything he found that was interesting and did not worry about organizing it. He used *CZNotes* as a memory aid, knowing that his notes would be available later when he went back to look at his collection of notes. P4 described using *CZNotes* as a form of external memory. The fact that participants were creating many *CZNotes* and linking them to the data would improve the chances that *CZNotes* would be returned in query results.

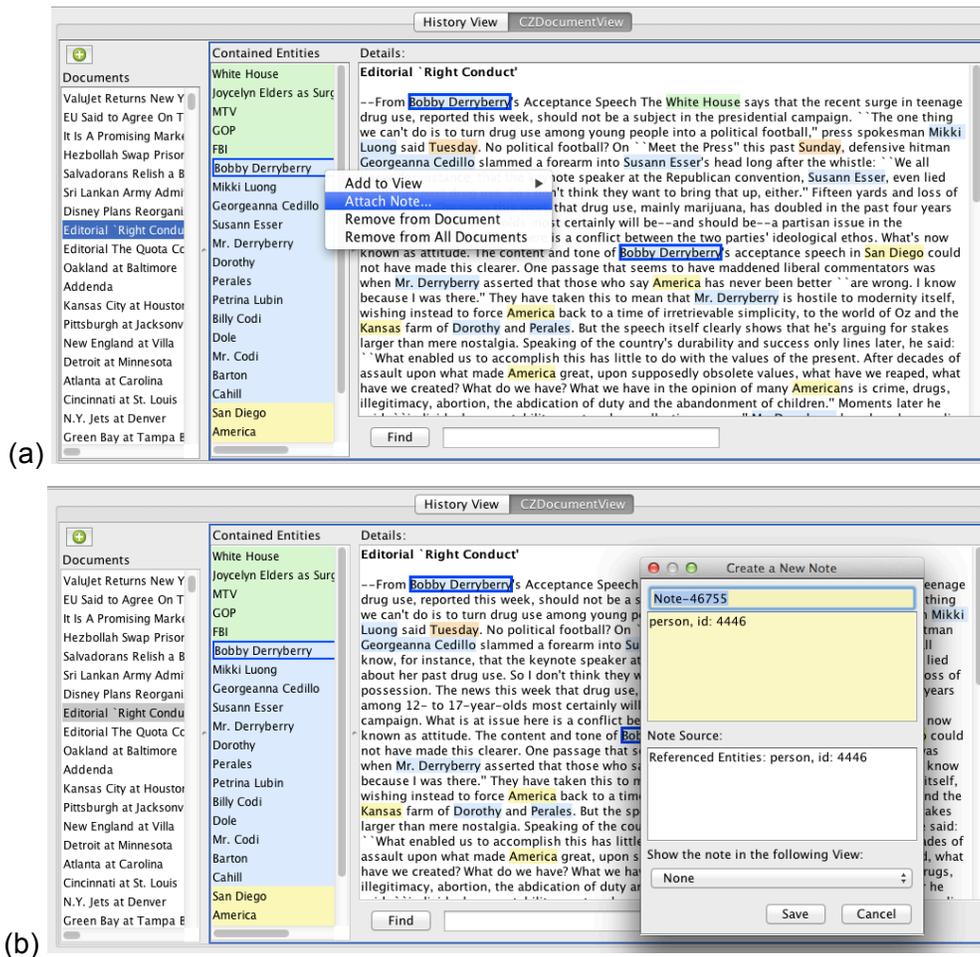


Figure 5-2. Attaching a CZNote. (a) P4 highlights text of entity “Bobby Derryberry” in document “Editorial ‘Right Conduct’”. (b) “Attach Note” action generates new CZNote referencing the entity Bobby Derryberry.

5.3.2. CZNotes Support for Continued Analysis

When continuing analysis from a previous session, participants P1, P3, and P4 stated that they did retrieve existing CZNotes created in previous sessions. One way for a participant to retrieve CZNotes was to add all CZNotes into a data view, or select previously created CZNotes from the *Note Repository* side panel and add them to a data view. Another way was to work with the CZNotes directly from the *Note Repository* panel.

1. P4 said, "... In a view I would say I want all the notes..."
2. Participant P1 added them to the *Hybrid Graph View* from the Note Repository side panel.
3. Participant P1 accessed *CZNotes* using the Note Repository side panel.

CZNotes were frequently used as a shortcut to retrieve data referenced in the *CZNote*. P3, for example, said he would sometimes use notes as a big pile of links, as a shortcut, to access all the documents related to a *CZNote*. There were multiple ways to do this but the following steps capture the general usage scenario:

1. Open relevant *CZNotes* in a *Data View*.
2. Retrieve the referenced entities or documents from the note and add them to the data views.
3. Continuing reading the documents (e.g. in the *Document View*).

P1 used his own *CZNotes* as shortcuts to the data. One of the first things P1 did when starting a new analysis session was to open and read the documents referenced in his *CZNotes*. P1 also shared his *CZNotes* with P3, who similarly used the steps above to access the data referenced in P1's *CZNotes*. A participant described that by using *CZNotes* as shortcuts, he avoided the need to re-do previous queries to retrieve data, i.e. *CZNotes* were used for retrieving data instead. P1 and P3 would then continue reading the data documents retrieved from the *CZNotes*.

We also wanted to know if participants used *CZNotes* as a way to query for data related to the note. More specifically, in *CZSaw's Hybrid Graph View*, double-clicking a graph node would find entities and documents related through bibliographic coupling and co-citation to the clicked node (See "*CZNotes in the Hybrid Graph View*" in Chapter 3). Therefore, double clicking on a *CZNote* node would also retrieve and display the documents and entities related to the note. In the *CZNote* log, we recorded the entity

types from which queries were made and the entity types returned to observe if participants made queries from *CZNotes* in the *Hybrid Graph View* (Table 5-4).

Participant 1		Query from Type	
		Non-CZNote	CZNote
Return	Non-CZNote	9	*
Type	CZNote	0	*

Participant 2		Query from Type	
		Non-CZNote	CZNote
Return	Non-CZNote	1	0
Type	CZNote	0	0

Participant 3		Query from Type	
		Non-CZNote	CZNote
Return	Non-CZNote	15	1
Type	CZNote	0	0

Participant 4		Query from Type	
		Non-CZNote	CZNote
Return	Non-CZNote	6	0
Type	CZNote	0	0

Table 5-4. Number of times data or CZNotes were queried using the get-related function in the Hybrid Graph View. The * represents P1 querying from a CZNote, however, no return type was logged perhaps because CZSaw “froze” and never completed the query.

For example, the CZNote log indicated P1 attempted to find entities related to his CZNotes within the *Hybrid Graph View*. When re-executing his script, we noticed that he appeared to want to find related entities to multiple CZNotes at the same time. However, the system froze when we attempted to execute the query (the system was still processing the query after 10 minutes). Therefore we do not know what type of results was returned. Nevertheless, it was interesting that P1 used the “find-related” query on a

CZNote. In another script, we successfully re-executed P1's script and observed the usage scenario shown in Figure 5-3.

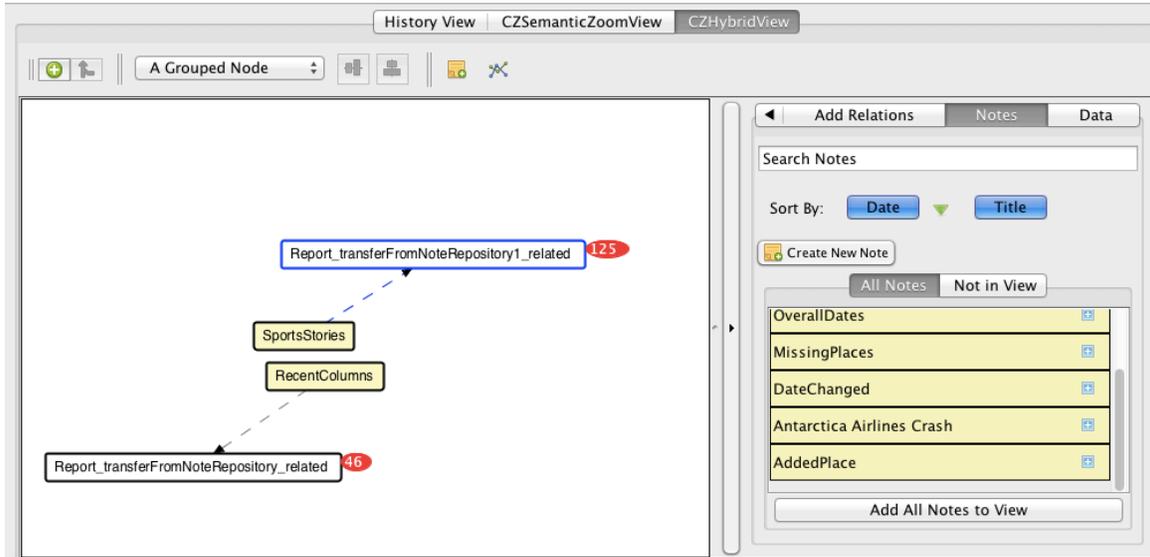


Figure 5-3. CZNote usage scenario - Opening the CZNote Repository to retrieve 2 existing CZNotes (Yellow nodes) and subsequently retrieve entities referenced within the 2 CZNotes (White nodes).

P1 added two CZNotes he had previously created, to the *Hybrid Graph View* and subsequently used them to retrieve the related documents (by double clicking on the CZNote nodes *SportsStories* and *RecentColumns*). P1 stated that he took advantage of the brushing and linking in CZSaw by selecting these returned documents (the white nodes in Figure 5-3), which then highlighted these same documents in the SZV. These documents were then filtered out of the SZV. This is further described below in Section 5.3.5, *Live Links*.

The usage of CZNotes to trigger further data queries indicates a potential for analysts using CZNotes as an aid in identifying next steps in their analyses. We believe this is a promising area for more in depth research in the future.

5.3.3. Reviewing CZNotes

Participants also described that they reviewed *CZNotes* explicitly in several scenarios:

1. P1 reviewed his *CZNotes* when starting a new analysis session in order to remember what notes he had and also to summarize his notes in order to make a wiki post.
2. P3 said his review of *CZNotes* was “totally organic” and that he didn’t have a real pattern of notes review.
3. P1 reviewed his *CZNotes* in order to create a summary post for the wiki.
4. Participants also reviewed each other’s *CZNotes* when notes were shared through the script. For example, P3 reviewed P1’s notes when P1 proposed a solution to the analysis.
5. *CZNotes* were reviewed by P1 and P3 in order to write the final report.

5.3.4. Building a Story with CZNotes

One way participants used *CZNotes* was to create notes as a way to store small chunks of information during the analysis. Usually at some point however, they wanted to gather up all these disparate pieces of information and aggregate them. P4 said this was analogous to gathering all the documents one would need for a writing project. P3 said this was similar to organizing a set of sticky notes to compile something coherent. Ultimately, the users wanted to build a story from the data. Participants suggested that combining relevant *CZNotes* could help to build up a narrative or a story of what was happening in the data. They reviewed their *CZNotes* to create a summary to post on the wiki. P1 took this compilation approach by creating one large *CZNote* containing the summary of the whole story. This *CZNote* referenced other notes that it summarized. For example, P1 created a *CZNote* about the seizures of weapons that referenced a *CZNote*

about the weapons being stolen. Similarly, P1 used his *CZNotes* to support the writing of the final report.

It seems that reviewing and summarizing *CZNotes* was important when the analyst has gathered a large number of notes. However, in P4's case, he said the number of *CZNotes* he collected was small enough to remember most of them and was able to summarize and post his findings to the wiki without either reviewing the notes he created or creating a summary *CZNote*. We could argue that, no matter the size of one's set of notes, reviewing past notes would help to ensure that nothing is missed.

5.3.5. *Live Links*

The connection between the *CZNotes* and the data is the key capability enabling the retrieval of data through the notes and the support of continued analysis using previously created *CZNotes*. All participants discussed the desirability of this live link between *CZNotes* and the data. P3 said it reminded him of hypertext links. P1 said this linkage between *CZNote* and data cannot be done with a wiki. These *live links* were not only useful when working with a participant's own *CZNotes* but also to retrieve data from *CZNotes* of another team member. P3 collaborated with P1 to access the data linked to P1's set of shared *CZNotes*. P3 stated that without these linkages from the *CZNotes* to the data, *CZNotes* would simply be a "big old box of words". Furthermore, P1 said you might as well create the notes outside of *CZSaw* if there were no such *live links*.

P3 described the *live links* feature as an important functionality to keep even during the final production process. He said, "When you're doing final production you need the ability to work with an image, a screenshot. You need to be able to do styling, basic text formatting. So, all your standard word processing stuff. But it would be best if those features were in conjunction with all the power of *CZNote* provides - the ability to get links and stuff like that."

Interestingly, *live links* were used also as a filtering tool by P1. P1 began collecting the (irrelevant) sports score articles into one *CZNote* by adding them as references to the note. When working with the *Semantic Zoom View* to analyze the

entire data set, he wanted to filter out irrelevant articles such as sports scores. While the SZV had a filter function, there was no easy method to select all the irrelevant articles. Therefore, he opened the *CZNote* with the irrelevant articles in a *Hybrid Graph View* and selected all the irrelevant articles contained within. This selection of the articles also triggered a selection in the SZV. With the irrelevant articles now selected in the SZV, he could use the SZV's "filter" functionality to remove them from the view and continue with his analysis.

5.4. CZNote Reminders

When iteratively performing data queries, it might be beneficial to be reminded of previous analytical findings by using *CZNotes*. Thus, we were interested to see the effect of including *CZNotes* in data queries. By default, all entity types are searched, including the notes. Would participants be aware that their *CZNotes* are being returned in search results and would this remind them of the past *CZNotes* they created?

Search Results	Count
Non-CZNotes in Search Panel	104
Mixed in Search Panel	24
CZNotes-Only in Search Panel	34
Total	162

Table 5-5. Number of searches that returned Non-CZNotes only (Any entity type other than CZNotes), CZNotes-Only, and Mixed entity types.

The log file data showed that 36% of the 162 queries (34 + 24 out of 162) made by all participants using the search panel returned *CZNotes*, 15% (24 out of 162) contained a mixture of *CZNotes* and other Entity Types, and 21% (34 out of 162) contained only *CZNotes* (Table 5-5). This shows that *CZNotes* were, in fact, being "fed back" into subsequent query results to potentially support iterative analysis. If *CZNotes* were shared among collaborators, finding them in query results could also reveal what notes others have made on the query topic.

We asked participants if they noticed *CZNotes* being returned as part of the results of a query. P3 felt the returned search results as “duplicating” the data. When this participant saw that *CZNotes* were shown in the results, he would say, “... oops, I forgot to turn off the notes showing up”. When asked if *CZNotes* in search results were distracting, both P3 and P4 stated it did not upset them as they said they simply mentally disregarded them. P1 said he filtered his search results to only include the new article reports, thus did not see *CZNotes* in his search results. However, he would search for notes in the Notes repository.

Feedback from P4 mentioned that seeing his *CZNotes* in a search result reminded him that he had gone down *this* path before and not to pursue it any further. P4 commented he would want notes to provide context. He said, “I would want to keep something peripheral though. Just some sort of clue that notes was [sic] there”. This scenario shows the potential benefit from feeding *CZNotes* back into subsequent query results.

Participants did hypothesize that such reminders, i.e., having *CZNotes* returned in search results, would be helpful in collaborative situations. For example, seeing another collaborator’s notes about “The President” being returned in a query would show that someone else was looking into “The President” as well. However, participants did not indicate that this collaboration scenario occurred during this user study.

From this feedback, we see that including notes in search results can provide reminders of previous analysis paths as in P4’s case described above. All participants further suggested that an analysis that went on for a longer duration, i.e., more than the 3 weeks of the study, would benefit with note “reminders”. They also suggested that such *CZNote* “reminders” would be useful in cases where data is constantly changing, requiring users to re-do the same queries on the data.

5.5. Using *CZNotes* to Create New Relationships

At the time of this study, *CZSaw* only supported bibliographic coupling, co-citation, and containment relationships. Two entities contained in the same document are considered related by co-citation. A *CZNote* containing two entities would also create such relationships among the entities. Therefore, we wanted to know if participants would use *CZNotes* and its entity referencing capability to create new relationships among the data that did not originally exist. For example, a user might create a *CZNote* about two entities they felt should be related, e.g., “Bob” was seen in “Canada”. If “Bob” and “Canada” were not in the same document in the initial data, then “Bob” and “Canada” were not considered to be related. The new *CZNote* referencing “Bob” and “Canada” would create such a relationship.

All participants were aware of the bibliographic and co-citation relationships in *CZSaw*. When participants were asked if they used *CZNotes* to create new relationships, P1, P2 and P3 responded that they may have created such relationships as an outcome of adding their analysis notes, but they never did so intentionally through bibliographic and co-citation (P4 was not available for this particular discussion). The *CZNote* logs from participants P1 and P3 indicated that indeed such new relationships were created. This relationship creation was detected during our analysis of the *CZNote* log created by instrumenting the *CZSaw* code.

Participant	Number of times Entities were Added to a <i>CZNote</i>	New Relationship Among Entities Created
P1	88	9
P2	0	0
P3	5	4
P4	4	0

*Table 5-6. Number of times each participant referenced an entity within a *CZNote* and number of new relationships created by adding such references to *CZNotes* that did not exist in the original data set.*

Table 5-6 indicates a total of 13 new relationships between entities were created by adding references to *CZNotes*. The use of *CZNotes* enabled participants to build new connections among the data. Although participants did not intentionally create bibliographic or co-citation relationships among the data using notes, the log file data revealed that two participants, P1 and P3, did create such relationships implicitly. This indicates the potential benefit of feeding new information created by the *CZNotes* (i.e., new relationships among the data) back into the data, which will be included in future data queries. Such new relationships feed back into the data and will support future queries and analysis of the data, as described in Section 5.3, “Support for Continued Analysis”. Entities that an analyst considers to be the same person, e.g. Entity A and Entity B, can be referenced in the same note. Because such entities are co-cited within a single note, they have a relationship. Thus in the future, when a query is made for Entity A to find related entities, Entity B will be returned in the results.

P1 hypothesized that such newly created relationships feeding back into a “central knowledge base” has potential benefit in a “long term intelligence situation”, however, due to the “short” duration of this study and the fictional data set, participants did not observe any obvious benefits of creating these new connections during the study. P1 said “if this was a longer analysis and there were more things happening, I would want that connection to show up later for me to remember that - oh you know, I said these things are connected.”

Furthermore, all participants agreed that relationships created with *CZNotes* should be visually distinct from the relationships found in the original data set. In this version of *CZSaw*, the *Hybrid Graph View* uses a dashed line to represent the links between *CZNotes* and other data entities, while all other links are represented with a solid line.

Another finding was the use of *CZNotes* to implicitly “merge” entities, rather than using entity merge functionality built into *CZSaw*. When entities are merged using *CZSaw*’s entity refinement tool, a new entity is created with a new entity identifier. If this newly merged entity is merged with another user’s merged entities, there may be a conflict of identifiers. Therefore, P1 used *CZNotes* to link entities together by referencing

them in a single *CZNote* rather than using the *CZSaw* entity merge functionality. *CZNotes* can link entities together by creating a co-citation/bibliographic relationship among the entities.

5.6. Add New Data into CZSaw

Participants had the ability to add new data (i.e., data not in the original data set) into *CZSaw* by including this new data in a *CZNote*. For example, the analyst may have found a relevant news article online that she wanted to bring into *CZSaw*. She could copy the text into a new *CZNote* and add references to existing entities in the initial data set, or she could extract new entities from the new data. In this way, she is bringing structure to the new data and making this new information searchable within *CZSaw* and its data views.

In this study, however, participants did not extract new entities (i.e., bring in external data) within *CZNotes*; the reason for this, they said, was that external information was not relevant for the analysis of a fictional data set used in this study (i.e., the VAST Challenge data). Confirming this, the *CZNote* logs showed no indication participants extracted new entities within *CZNotes*. All entities extracted from *CZNotes* were existing entities. Only P4 extracted new entities, but from a data document that was missed certain entities during initial import by the Named Entity Recognition tool, not from a *CZNote*. This area would require further study, where the import of external data through *CZNotes* could supplement the original set of data.

5.7. Visualizing CZNotes In Data Views

We were also interested in how *CZNotes* would be used in a VA tool, *CZSaw*, which allowed both the data (i.e., *CZSaw* documents) and the synthesized knowledge (i.e., *CZNotes*) to be visualized in the same space.

	DV		HGV		SZV		Search Panel “Add to View”
	D & D *	Context Menu	D & D *	Context Menu	D & D *	Context Menu	
Non- CZNotes	11	36	11	66	1	4	83
CZNotes	1	0	32	14	1	0	0
Mix	0	0	0	0	0	0	0

Table 5-7. This table shows the frequency of the different ways CZNotes and Non-CZNotes were added to the CZSaw data views: Hybrid Graph View (HGV), Semantic Zoom View (SZV), and Document View (DV). CZNotes were primarily visualized in the Hybrid Graph View. * D & D = Drag and Drop.

Table 5-7 indicates that CZNotes were added to the Hybrid Graph View 46 times using the two methods of drag and drop (n=32) and a context menu (n=14). Drag and drop means CZNotes were dragged from another data view into the Hybrid Graph View. A user can also right click on a set of selected CZNotes to trigger a context menu to pop up with an option to send the CZNotes to another data view. We also discovered through participant discussions that P4 added CZNotes to the Document View using the “add all” feature. As we can see from

Unfortunately, the log file only records the set of entity types added to a data view at the time the “add to view” action was invoked by the user. The log did not capture whether or not data views contained a mix of data types (i.e. CZNotes and other data entity types). From the log, we see that CZNotes were added to the different Data Views, especially the Hybrid Graph View. However, from replaying a small number of participant project scripts, we did observe that CZNotes and other entity types were sometimes used in the same data views. For example, when participants wanted to retrieve data from a CZNote, they would add the CZNotes to a HGV and retrieve all the documents referenced in the note into the same HGV for further analysis of these documents.

We were also interested in how *CZNotes* were organized and laid out in the data views. P3 commented that once he started profiling the entities within the data, he wanted to layout his *CZNotes* in the *Hybrid Graph View* in order to see all the notes at once. P3 said, “I do [review notes]. I tend to put them in a *Hybrid Graph View* because I want to be able to see them all. Actually ... once we started the profile I started spatially organizing my notes in the *Hybrid Graph View*. And grouping different things and that kind of stuff. Because I needed to be able to see it ... them all in a spatial layout, not just in that list view. I need to be able to put ... hey, these are these, these are those. President here”. P4 had a similar desire to organize his notes to prepare them, for example, like a writing project.

However, P1 described some hindrances related to laying out *CZNotes*. One issue was that the layouts of views were difficult to re-open due to *CZSaw* performance issues when re-running long scripts. P1 would start a new analysis session using a script with only the data commands saved from a previous session (e.g., *CZNote* creation and edit commands) without the view related commands (e.g., adding *CZNotes* to a data view and then positioning the *CZNotes*). A further difficulty in organizing *CZNotes* visually was the limited size of the *HGV* on the participants’ laptop machines, which caused *CZNotes* to overlap one another. P1 said, “I have not really organized my notes. That is, moved them around in a *Hybrid Graph View*, because [for] each session I create a new project and just use the data commands from the previous ones. Thus, any visual organization of my notes would not be maintained anyway.”

The flexibility of having *CZNotes* added to Data views, however, enabled participants to interact with *CZNotes* in interesting and unexpected ways. The *Hybrid Graph View* was used as an environment to facilitate adding entities to *CZNotes*. Due to drawbacks in this first design of *CZNotes*, it was not easy to reference multiple entities in an existing *CZNote*. For example, one cannot place a relevant entity or document into an existing *CZNote* to be referenced, but rather, must extract these relevant entities within the text of the note or use the *HGV*’s *CZNote* node’s “add references” feature (see chapter 3) to add these entity references. P1 developed a process to use this add-references feature in *HGV* to add references to data to existing *CZNotes*. P1 dedicated one entire *HGV* to one or two *CZNotes*. When he found documents or entities he wanted

to reference in one of the *CZNotes* in the dedicated *HGV*, he would add these entities to the same *HGV*. He would then drag the *add-reference* icon on to the entities he wanted to reference. Thus, P1 came up with this work-around to “more easily” add entities to an existing *CZNote*.

P1 also used the *Document View* to review his *CZNotes* because it provided a larger reading area since other data views and the note repository were limited in size for viewing *CZNotes*.

CZNotes were seldom used in *SZV*. From participants’ feedback, it seems the data set used in this study was small enough that the participants claimed they did not need to visualize *CZNotes* in the data views but could recall notes from memory.

Due to design and performance issues in the current version of *CZSaw*, we could not determine if visualizing *CZNotes* in a data view would provide any benefits--an area for further research.

5.8. Collaboration and Sharing *CZNotes*

Collaboration support was not a primary goal in *CZNote*’s design. Despite this lack of designed-in support, participants used various workarounds to share *CZNotes* with one another by (1) sharing, via a wiki, the script file containing the note creation commands and/or (2) posting notes’ content directly on the wiki.

It was clear from their comments and actions that participants wished to share analytical artifacts despite the current barriers to doing so in *CZSaw* and *CZNotes*. *CZSaw* does not provide a centrally shared online repository of data and *CZNotes*. However, all user interactions are recorded in a script file and this file can be given to other team members. This shared script file can be re-executed within any *CZSaw* instance and will import any *CZNotes* created by collaborators. Unfortunately, since there is currently no automated way of merging two scripts into one, collaborative activities require manual merging and managing of various scripts from collaborators.

5.8.1. Wiki Versus CZNotes

At the beginning of the study, participants initiated the use a *CZSaw* research group wiki to share their analytical findings during the course of their analysis. They foresaw the need to share their findings in a central and easily accessible location and understood the limitations of *CZSaw* and *CZNotes* collaboration support. We encouraged, but did not require, that participants first write any notes in a *CZNote* before posting them onto the wiki. We did not analyze the contents of the wiki posts the participants made, as this wiki data was not in the original scope of our study design.

Participants also anticipated that wiki postings could be used for different purposes compared to *CZNotes*. One difference was that *CZNotes* are initial findings and collections of synthesized knowledge obtained during data exploration, whereas the wiki posts might be a distillation of multiple *CZNotes*. P3 said, “Well, I think the wiki is going to have a different function. It's going to be a transferring of distilled ideas already. Whereas the *CZNotes* are really about what you're doing and how you're working. So the next step after you work with notes is the wiki”. P3 describes *CZNotes* as lightweight and fast way to capture information in as few words as possible whereas a wiki would be used for long drawn out discussions. Furthermore, P3 said, “[unintelligible] they're going in notes. What's going on the wiki is distillation of our previous thoughts. You'll see reflection of the note in the wiki. Because all the work is being done with thoughts to whatever and then its ... okay, here's the final conclusion. And the only reason we're going to the wiki at that point is because there's no way of easily sharing information with your friends from within the notes”.

Participants' feedback indicated wiki postings were in fact a distillation of *CZNotes*. Most participants would create a summary of their *CZNotes*, rather than posting all their *CZNotes*, and share the summary on the wiki. P1 said he created a “super-note to summarize what [he] found and described the documents in order by dates. That note's content is basically copied to the wiki page exactly as it appears in the actual note.” He would also share the script file to recreate that note on the wiki. P3

noted, “At some point we need to be able to share these notes in their *CZSaw* format so that we can look at them and distill back to the wiki”.

P4 also created wiki entries summarizing some of his analytical findings, but this summary was done from memory because the set of notes he created was small enough to do so and therefore he did not refer to the *CZNotes* he created in his wiki postings.

5.8.2. Pursuing Different Threads

All the participants agreed to use a divide and conquer approach to pursue different threads of analyses as they emerged. They used a portion of the focus group session time to discuss possible threads and assign them to team members. Participants created one page per thread on the wiki to share their findings with others. The wiki offered a way for participants to setup different pages for each of these threads. They discussed that if one participant found information about another participant's thread; they could post on that thread's wiki page. We noticed a similar behaviour in using *CZNotes* individually as a categorizing tool (see “*CZNotes* as a Categorizing Tool” in section 5.9).

5.8.3. Live Links and their Need

While the Wiki had the advantage of being a central share of information and easily accessible, it did not provide the “live link” capabilities of *CZNotes*. Participants felt *CZNotes* had the advantage of linking to data and other notes as described in section 5.3.5. P1 said, “... the wiki's not live. It won't live connect you to the entities in *CZSaw*, but like I've just started a new project last night and so I have notes in two different places now. So I would need to manually put those scripts together if I wanted to keep all the notes. So I'll probably transfer the notes from both of these two sets of projects to the wiki and be able to see them in there. Yeah you lose stuff because you can't live link back but ...”. P3 said that “ The ability to add references to notes were the most used feature in my analysis. I used notes to reference documents of importance.” The ability to use *CZNotes* as a link to the data was important enough for the participant's to spend

time to export their notes into a script file and store on the wiki to share with the rest of the team.

5.8.4. Sharing CZNote Scripts

Participants encouraged one another to upload *CZNote* script files to the wiki along with the notes posted on the wiki. The desire to share the script file containing the *CZNote* creation commands was so that collaborating participants could use the *CZNotes* live in their own running instance of *CZSaw*.

To utilize the live links of *CZNotes*, P1 shared his script containing only *CZNote* creation commands with P3 (by uploading the script to the wiki) for P3 to download into his *CZSaw*. P1 was the only one who shared his *CZNotes* with others through the *CZSaw* script. He primarily wanted to share only the *CZNote* creation commands in his script rather than all his recorded actions. Over time he refined his method of extracting *CZNotes* from the *CZSaw* script to share with others and also to use for himself in subsequent *CZSaw* sessions. Initially, he used *CZSaw*'s "extract data commands" option to filter the script to only commands that manipulate the data, e.g. extracting entities and also creating and editing *CZNotes*. However, he decided that he did not need the edit *CZNote* commands, as he only wanted the final state of the note. Also, re-executing these *edit-CZNote* commands increased the time it took to load the script. Therefore, he manually changed the last edit *CZNote* command of each note to a create *CZNote* command and deleted all preceding edit *CZNote* commands. Thus, the final state of each *CZNote* was saved in refined scripts as a single create note command. P1 described this process: "the last thing I did was edit the text of the two notes that I wanted to keep, and then I closed *CZSaw* and I opened up the script in a text editor and just grabbed the last couple of transactions and changed them from editNote to the note constructor command". In essence, P1 simply wanted an easier way to just output the final state of his *CZNotes* rather than go through this cumbersome process.

P2 and P4 did not describe using the script to share *CZNotes* but they did use the wiki to share some of their analytical findings and some content from their *CZNotes*. P2 did note however the need to have a central repository for sharing *CZNotes* that could

be used within *CZSaw*, saying "... I want to be able to use that [shared note] in *CZSaw*. In my own version of *CZSaw*. *CZSaw* should be able to do it."

5.8.5. Using Other Participant's CZNotes

When discussing collaboration with *CZNotes*, participants desired to see other's *CZNotes* and also access the references of the *CZNotes* within their own *CZSaw*. Unfortunately, this is not possible with the current version of *CZSaw* as there is no online repository of *CZNotes* that *CZSaw* users can connect to, to share *CZNotes*. However, scripts containing *CZNote* commands were shared in this user study. P3 described his process of importing P1's *CZNotes* into his own *CZSaw* session and merged P1's set of notes with his own. P3 would add P1's *CZNotes* into the *HGV* and expand the notes to retrieve the documents the *CZNotes* referenced. Then he would drag these documents to the *Document View* for detailed reading. In this way, P3 used P1's notes to review documents that P1 hypothesized were part of the solution to the analysis task. We similarly observed *CZNotes* used as a shortcut to the data during individual analysis by P1 for reviewing and continuing analysis from his own *CZNotes* in subsequent analysis sessions.

When we asked P3 whether he edited P1's notes, he said he did not remember. He commented that in this particular situation, he did not find it necessary to know whose *CZNotes* were whose.

P3 said, "I think I did [edit P1's *CZNotes*]. I'm not sure. I definitely looked at them. I can't tell you off hand. And I never really thought ... oh this is [P1]'s note, [unintelligible] okay I've got a whole bunch of more notes. Because I also added all my notes from the past that I was bringing in too. So it just became one big fat collection of notes."

Participants, however, did want a visual indicator to show the author of a *CZNote*. This information was hidden in a tab within a *CZNote* that required some work to access. A well-known example of this would be Microsoft Word's '*track changes*' feature where each comment is color-coded and tagged with the commenter's name.

5.8.6. Private Versus Public

We observed from this study that the content of a *CZNote* might be influenced depending on the visibility and the location of where the notes are shared. For example, there is a qualitative difference in note content for a wiki and for a *CZNote*. When collaborating, whether or not *CZNotes* are shared via script or wiki, participants only wanted to share notes they felt were important and not every note they created. Participants explained that they did not want to share everything they found. If *CZNotes* were automatically shared with participants, then participants may have been more conservative in the notes they took. Similarly, if participants only had access to the wiki for sharing notes, we hypothesize that they would be much more selective in posting (i.e. capturing notes) as it would be available for everyone to see. *CZNotes* were seen to be private unless the *CZNote* script is shared, in which case, the participant would filter the script to include only what was “relevant”. P3 said “... one of the ways we can collaborate here is we can start sharing specific notes.” P1 did the aforementioned process of narrowing down the set of *CZNotes* he wanted to share in the script using a text editor before uploading it to the wiki. Similarly, notes were also filtered manually when transferring *CZNotes* content to the wiki. P1 only transferred what he wanted to share with others.

This suggests the importance of marking and making notes as either public or private. The wiki is a sharing of distilled ideas as described by participants earlier, which is different from the use for *CZNotes*. P3 describes the use of the wiki is preceded by the use of *CZNotes* (The wiki is to store the distilled knowledge from *CZNotes*). This distilled knowledge is what the participants wanted to share rather than all the bits and pieces they captured in *CZNotes*. Even when posed with the scenario of an easy mechanism to share *CZNotes* with *CZSaw*, P4 said he would not share notes with others unless the notes represented some significant finding.

The participants also took time to determine what to share on the wiki. In a way, participants are providing their insight and decision-making in this process. They are making a decision on what is deemed important enough to share i.e. post on the wiki or include in the script. P4 said that he would not post everything he took note of.

Therefore, in the future development of *CZNotes*, we need to consider private and public notes.

In summary, participants' use of *CZNotes*, both alone and collaboratively, shared many similarities. *CZNotes* were used to collect information, which were used as a basis to continue analysis in future analysis sessions. We see that this is also the case when *CZNotes* are shared. Collaborating analysts also used the shared *CZNotes* as a starting point for analysis. The difference is that prior to sharing *CZNotes*, they were pre-filtered to remove any *CZNotes* that were deemed unimportant. This filtering prior to sharing findings is also evident in the participants selectively posting to the Wiki.

5.9. *CZNotes* as a Categorizing Tool

Participants also described using *CZNotes* as a “shoebox”, i.e., as a container for categorizing interesting data found during analysis. Using *CZNote*'s entity referencing capabilities, participants were able, for example, to collect all “noise” documents into a single *CZNote*. P1 found many documents about sports scores, which he deemed irrelevant to the analysis and created a *CZNote* as a container for these “noise” or irrelevant documents. When subsequently analyzing the entire document set in SZV, P1 used this *CZNote* to filter out these “noise” documents from the data view by using the brushing and linking technique described in section above 5.3.5 above. Similarly, P3 described his need for a “shoebox” as “You just want a big box you throw things at.” In this case, it is multiple shoeboxes, one for each category of information to collect.

Another perspective is to think of *CZNotes* as “tagging” items put into them, rather than as a container (i.e., shoebox). Participants used the term tagging when describing the action of adding a reference to an entity or document in a *CZNote*. P3 commented that he used *CZNotes* as a way to tag documents or as folders. “I use [*CZNotes*] a lot as a tag. One word. Just a title. Slam some stuff in it”, said P3. P1 follows by saying “I do that a lot too.” P1 also explains, “... [my] use of notes was primarily as tags. One word. Add to note and then later on I'd go and look at those notes.”

The primary way participants used *CZNotes* to categorize information during this study was by thread of analysis. P4 described this working style of having one *CZNote* per thread of analysis and wanted to continually add more documents and information to each of them as he finds more relevant information. Similarly, P1 said, "... I was making a note for each of the suspicious activities or interesting things and then I made a super note to summarize what I found and described the documents in order by dates." For example, P1 created a *CZNote* referencing all the documents about the "Atlantic Airlines" from the data set. These were the summary *CZNotes* that were posted to the wiki. This type of use of *CZNotes* is very similar to the wiki thread pages described earlier. At times, he would name his *CZNotes* based on the search terms that he used when exploring the data.

Another way to think of a thread of analysis is the idea of profiling individual entities within the data. During one of the focus group sessions, P3 said, "I think we need to profile stuff in order to get a better picture [be]cause there is a lot of interconnectedness that we need to distil in to discrete categories. I think we can then use the notes things and we can create a single note for each profile or whatever and then use that to link onto the documents." They would create a note to profile the important entities in the data such as the person entity "The President". Each profile *CZNote* would have a reference to all the related data documents and various information about the entity.

The Semantic Zoom View also had a grouping feature that is specific to the view. However, P1 describes that he preferred to create such categorizations of the data with *CZNotes* and their references rather than with the SZV because he did not want the groupings to be tied to a specific view. The *CZNotes* were also much more easily exportable in the script as a single data command rather than a series of commands needed to generate the view with the grouping, which also takes longer to re-execute compared to the single create *CZNote* command.

Initially, *CZNotes* content contained mostly references to *CZSaw* documents with relatively few references to entities like people and places. However, P1 switched his

strategy from adding all the related documents of an investigation thread, e.g. The “Antarctic Airlines”, to simply adding the single “Organization” entity of “Antarctic Airlines”. P1 realized that documents could be retrieved through the entities contained in them. He noted that adding all the related documents may slow down performance because of the large collection of documents a single note references, whereas you have Antarctic Airlines as a single entity instead. Through this single “Antarctic Airlines”, he could find all documents that contain it. This shows the power of connections between entities.

5.10. Summary

The data collected in this preliminary study suggest that integrating *CZNotes* with *CZSaw* data was not only plausible but demonstrated several interesting uses of *CZNotes*. The following is a summary of the different *CZNotes* use cases identified by the performance of the participants in this study.

1. *CZNotes* were used as a *categorizing tool*. Relevant data and other *CZNotes* regarding a particular thread of analysis were collected within a single *CZNote*.
2. *CZNotes* served as a *shortcut to retrieve data* referenced in a note. This was done primarily at the beginning of a subsequent analysis session.
3. *CZNotes* were used as a bridge to *link existing data and synthesized knowledge* into a single knowledge element. *CZNotes* supported the creation of new relationships among the original data, as well as relationships between the data and other *CZNotes*.
4. *CZNotes* enabled users to link entities they thought were the same, i.e., provided a *way to merge or link entities*.
5. For one participant, seeing existing *CZNotes* in his search results served as a *reminder* that the search was done before and triggered him to pursue a different analysis path.

6. Discussion

To recap, the purpose of this study was to examine the possible benefits of integrating analyst's notes into the data set used by an entity-based visual analytics tool, *CZSaw*. Many existing visual analytic tools support note taking, but treat notes as separate entities within the tool and have a separate workspace to interact with them. We have taken a novel approach that integrates notes and data by representing notes the same way as the data in *CZSaw*. We were interested to see the effects of such an integration on the following research questions:

1. Will analysts be aware of their previous findings (notes created by themselves or others) during their analysis with *CZNotes* being returned in query results?
2. Will *CZNotes* be used as a starting point for continued analysis? For example, an analyst would first review *CZNotes* in follow-up analysis sessions.
3. Will *CZNotes* be used as a means to create relationships among the data through co-citation?
4. Will *CZNotes* be used as a way to bring external data into the system (*CZSaw*)?

To address these questions, we created a note taking capability for *CZSaw*, called *CZNotes*. Our design allowed notes to inherit the representation of *CZSaw*'s data documents. Therefore, *CZNotes* became part of the data set and could be visualized in *CZSaw*'s data views and queried with its entire existing data search mechanisms. This enhanced tool was evaluated in a preliminary study over 3 weeks by 4 participants working as a team to solve a complex problem.

The results of the study suggest that *CZNotes* did **support iterative analysis**. Participants found *CZNotes*' ability to link to the data to be useful. This linkage enabled users to retrieve data that were linked to a *CZNote*. Analysts used *CZNotes* as a starting point in their subsequent analysis sessions, often retrieving a set of linked documents in a *CZNote* for further reading and analysis. Participants described this use of *CZNotes* as providing shortcuts to the data. Also, participants wanted to use ***CZNotes* as a way to categorize findings**. For example, each interesting entity was profiled in a separate *CZNote*. The use of *CZNotes* as shortcuts and as a categorizing tool was unexpected. Participants often dedicated one *CZNote* for a certain topic or a profiling of an entity.

***CZNotes* could be seen in query results** and, for one participant, finding notes in these results unexpectedly **reminded** him that his current query was linked to a previous dead end analysis path. This prompted him to pursue another analysis path. Participants hypothesized that feeding *CZNotes* back into the data set could potentially be more useful if the duration of the analysis was longer (i.e., more than 3 weeks). P3 said, "On a long term analysis, it is essential to have notes show up as part of the search results. However, I did not experience a direct benefit of this feature because the analysis was so short. I had a basic understanding of what notes I had created and what was in them. If the scope was large, I absolutely would have needed them to be in the search results because I would have forgotten they existed."

CZNote log data showed *CZNotes* were used to create **new relationships among data entities** by adding references to entities in *CZNotes*. Participants did not report that they intentionally used *CZNotes* to create these new relationships, but some new relationships were created by the natural use of *CZNotes* to capture synthesized knowledge during analysis. This demonstrated that *CZNotes* could support the expansion of the network of data entities and also among data and synthesized knowledge. The analyst can build up this knowledge set (data + notes) explicitly within *CZSaw*. This can be beneficial in supporting the iterative analysis described above, as future browsing of data would take into account any such new relationships among the data. Not only is this beneficial to a single analyst, but also can be beneficial to a collaborating team. However, the current *CZSaw* and *CZNotes* system has limited support for collaborative work. Even so, some participants made the effort to share their

CZSaw scripts containing their *CZNotes* by allowing another participant to import their script.

We also wanted to see if *CZNotes* could be a **“vehicle” to bring in external information** by incorporating new information into *CZNotes* and then extracting entities from this information. Since *CZNotes* are integrated with the data, the user effectively adds new “data” to the data set. However, the session logs from the study did not indicate that this occurred. Participants explained that because of the nature of the fictional data set, they did not see the need to bring in such external data. It might be interesting to study the use of *CZNotes* with real world data that requires bringing in external data, such as information found in Internet queries.

***CZNotes* were added to CZSaw Data Views** and in some cases, *CZNotes* and other entity types were used in the same *Data View*. Some participants used the spatial layout capabilities of these views to layout their *CZNotes*. We observed participants added *CZNotes* to data views, **but not as a result of data foraging as we expected**. They primarily placed *CZNotes* into data views for the primary purpose using the interactions available within the view to retrieve the data referenced in the note such as “getting the related” entities of a note in the *Hybrid Graph View*.

6.1. Limitations

User interactions with *CZNotes* were not consistent across data views in CZSaw. In the current design, *CZNotes* were only editable in the *CZSaw Hybrid Graph View* and in the notes repository. For consistency, *CZNotes* should be editable in all *Data Views* in a similar manner. One participant requested that *CZNotes* be editable in the *DV* in future. This limitation prevented participants from fully utilizing *CZNotes*.

In future studies, more time should be provided for the participants to learn how to use both CZSaw and *CZNotes*. P1 was most knowledgeable of *CZNotes* features and thus his participation was of great value in the present study.

Due to constraints of time and resources, only a single coder analyzed the qualitative data. It would have been preferred to have at least one other investigator analyse the transcripts and journal entries to develop the set of codes used to analyse the session transcripts and journals. Consequently, this user study should be considered preliminary and a more detailed study is needed to solidify results. However, useful user information was obtained to guide the continuing development of the *CZNotes* capability within *CZSaw*.

7. Conclusion & Future Works

7.1. Conclusion

We introduced *CZSaw* and *CZNotes*, a visual text analytics tool that supports note taking for keeping track of analytical findings. *CZNotes* are designed to integrate with the data set and inherit the capabilities of documents *CZSaw* – text with extracted entities. In addition, *CZNotes* also has the ability to edit the text and refine the entity extractions. We wanted to investigate the potential benefit of representing notes in this way – as part of the data. *CZNotes* can also support progressive analysis in that future exploration and query of the data will retrieve and display previously created and related notes. *CZNotes* can also play a role in the synthesis stages of analysis as demonstrated by participants when they reviewed past notes to write their final analysis report.

Thus, we have created a note taking tool with the novel capability of representing notes *as data*. We found this approach is viable in a VA system that represents data as text documents that contain extracted entity objects, similar to Jigsaw. We ran a preliminary study, and observed a team of 4 participants interact with *CZNotes* in the *CZSaw* VA system.

We discovered unexpected usages of *CZNotes* including as a shortcut and as a categorizing tool. In terms of our research questions, we observed that *CZNotes* did support iterative analysis by showing up in search results, creating new relationships among the data and supporting report writing.

Furthermore, we observed *CZNotes* used in a flexible manner and participants described *CZNotes* as a flexible tool.

P3 said, “[a] note that you can link to something will always be the backup for any missing features. Folders. Because it can do all those things. It can encapsulate. It can link. It can do all this stuff. A note is a universal tool that you can solve other things your system is missing with. We’ve done that. We’ve used it as a folder. We’ve used it as these kinds of things. Because it’s so free form and flexible that it allows us to do those kinds of things. And every system needs to have something like that. You sometimes come across ... you need to do this ... so you put it in a note”.

This comment suggests that a VA tool should provide a means to freely capture information without forcing the user to adhere to a specific structure.

From designing *CZNotes* and running the user study, we have shown the possibility of integrating synthesized knowledge with a data set to potentially provide continuous feedback of synthesized knowledge (*CZNotes*) back into the knowledge set (data + notes) for subsequent analyses.

In conclusion, the detailed data results were mixed, in part due to the small number of participants and also to the relatively short time period over which data was collected. However, the data does clearly suggest *CZNotes* have significant potential value and this work should be extended. The *CZNotes* user interface should be updated to provide an improved user experience and interaction. Other minor improvements should be made to *CZNotes* and *CZSaw* followed by an extended study making full use of the results of this preliminary study. We describe our future directions below.

7.2. Future Directions

An important area to refine for *CZNotes* is a simpler and more intuitive user interface and experience. Participants found simple things like adding entities to an existing *CZNote* to be difficult. They also wanted *CZNotes* to be ubiquitous throughout the system – available and editable everywhere. Participants may have used notes more if there were more support for them in the other views (*Semantic Zoom View* and the *Document View*) such as the editing of the notes. A *Document View* was not intended to

be editable in the initial design, but P1 would have liked to be able to edit notes in the *Document View*.

We also need to design a suitable framework that enables collaboration with a central repository of *CZNotes* that is accessible by all users. The challenge is to incorporate the use of the *CZSaw* script into this collaborative framework.

When such a collaboration system is in place, we would also want to add two additional features to *CZNotes*. One is the ability to mark a note as public or private and the second is to provide a visual indicator of who the author of a *CZNote* is in order to distinguish an analyst's own notes compared to a collaborator's notes.

CZSaw performance, as a whole, needs to be improved. This certainly includes *CZNotes*. Participants did not use certain features of *CZNotes* because of these performance issues.

An issue encountered when sharing *CZNotes* is how to handle sharing an edited *CZNote* after the original *CZNote* has already been imported and used by another collaborator. If P3 had edited P1's *CZNote* "A", and P1 also edited *CZNote* "A", there was an issue of P1 sharing his *CZNote* with P3 again. Due to this reason, P1 did not update the *CZNote* script file he shared on the wiki because it could be out of sync with P3's. P1 said "So the script that I put up on the wiki ... I haven't updated it because it's kind of a pain to pick out all those ... just those specific notes from my script and make it a simple, just creation of note script. Plus if I updated, would that be that useful? Because [P3] has already used the original one, right? So we'd have to merge both our changes back together. So I just updated the text on the wiki but I didn't update the actual attached script file." Therefore, we also need to consider the issue of merging multiple versions of *CZNotes* together into one, whether it be *CZNotes* from a single analyst or merging *CZNotes* from multiple analysts. Perhaps a versioning or history mechanism for *CZNotes* should be considered.

Participant P4 brought up an issue of how to combine *CZNotes* into a final report at the end of the analyses. P4 said, "[s]o another thing is ... let's say people have been

trading these notes back and forth. [P3] has picked up a few and [P4]'s picked up a few from some other place and now you want to write this up. So everybody has their own different bunch of notes and so how do you get it all together in the end?" How to utilize CZNotes to create a final report would be an interesting topic to focus on in future studies.

This 3-week long study seemed too short to truly gauge the impact of *CZNotes*. We believe a longer study examining the following areas in greater depth would be worthwhile:

1. Using CZNotes to bring in external data – we also need to consider a real world data set where bringing in external data, for example, from internet searches, is applicable as participants did not do so in this study because of the fictional data set provided.
2. Investigate the effects of feeding back information generated in CZNotes back into the data set to benefit the analyst, as participants hypothesized, in subsequent queries as well as in collaboration where the data set is shared among multiple analysts.

7.2.1. *CZNotes and Hypertext*

When *CZNotes* were initially conceived, they were designed to have the same representation as *CZSaw Documents*, which was inspired by *Jigsaw*. The capabilities of a *CZNote* are (1) the text of a note can be edited and (2) references to entities can be added or removed. Participants thought of *CZNotes* as a structure similar to HyperText. We realized the network of CZNotes could somewhat be considered a miniature version of the World Wide Web contained within *CZSaw*, in that information documents are linked together. The Internet is filled with data and information (i.e. webpages), but there is also synthesized knowledge such as web pages containing opinions, summaries of information, hypotheses regarding the “data” or raw facts. This raises the question, how do we distinguish data and notes (i.e. synthesized knowledge)? This study helped us realize how the basis of and the flourishing of the Internet were, perhaps, formed.

Furthermore, P3 suggested that if the representation of *CZNotes* took on the standards of HyperText, it would not be unthinkable to be able to export *CZNotes* as a website that can then be used to share with others. This is also an area worth pursuing.

References

- ATLAS.ti Scientific Software Development GmbH. (2011). ATLAS.ti - Qualitative Data Analysis. Retrieved from <http://www.atlasti.com/>
- Bartas Technologies. (2011). Transcriva 2. Retrieved from <http://www.bartastechnologies.com/products/transcriva/>
- Bier, E. A., Card, S. K., & Bodnar, J. W. (2008). Entity-based collaboration tools for intelligence analysis. *2008 IEEE Symposium on Visual Analytics Science and Technology* (pp. 99–106). IEEE. doi:10.1109/VAST.2008.4677362
- Bier, E. A., Card, S. K., & Bodnar, J. W. (2010). Principles and tools for collaborative entity-based intelligence analysis. *Visualization and Computer ...*, 16(2), 178–91. doi:10.1109/TVCG.2009.104
- Bier, E. A., Ishak, E., & Chi, E. (2006). Entity Workspace: An Evidence File That Aids Memory, Inference, and Reading. *Intelligence and Security Informatics*, (1), 466–472. Retrieved from <http://www.springerlink.com/index/j701mm35046603w1.pdf>
- Canfield, C. M., & Sheffield, D. (2011). Interactive data analysis with nSpace2®. *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 327–328). IEEE. doi:10.1109/VAST.2011.6102497
- Chen, V., Dunsmuir, D., Alimadadi, S., Lee, E., Guenther, J., Dill, J., Qian, C., et al. (2010). Model based interactive analysis of interwoven, imprecise narratives: VAST 2010 mini challenge 1 award: Outstanding interaction model. *2010 IEEE Symposium on Visual Analytics Science and Technology* (pp. 275–276). IEEE. doi:10.1109/VAST.2010.5653060

- Chen, V., Dunsmuir, D., Kadivar, N., & Lee, E. (2008). Interactive Demo: Using CZSaw to Analyze Entities in Collections. *sfu.ca*, 1–2. Retrieved from <http://www.sfu.ca/~shaw/papers/Chen-Demo-VAST10.pdf>
- Chen, Y., Barlowe, S., & Yang, J. (2010). Click2Annotate: Automated Insight Externalization with rich semantics. *2010 IEEE Symposium on Visual Analytics Science and Technology* (pp. 155–162). IEEE. doi:10.1109/VAST.2010.5652885
- Chen, Y., Yang, J., Barlowe, S., & Jeong, D. H. (2010). Touch2Annotate: generating better annotations with less human effort on multi-touch interfaces. *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*. New York, New York, USA: ACM Press. doi:10.1145/1753846.1754042
- Cheng, W., & Gotz, D. (2009). Context-based page unit recommendation for web-based sensemaking tasks. *Proceedings of the 13th international conference on Intelligent user interfaces - IUI '09* (pp. 107–116). New York, New York, USA: ACM Press. doi:10.1145/1502650.1502668
- Dunsmuir, D., Baraghoush, M. Z., Chen, V., Joorabchi, M. E. M. E., Alimadadi, S., Lee, E., Dill, J., et al. (2010). CZSaw, IMAS & Tableau: Collaboration among teams: VAST 2010 Grand Challenge award: Excellent student team analysis. *2010 IEEE Symposium on Visual Analytics Science and Technology* (pp. 267–268). IEEE. doi:10.1109/VAST.2010.5653035
- Dunsmuir, D., Lee, E., Shaw, C. D., Stone, M., Woodbury, R., & Dill, J. (2012). A Focus + Context Technique for Visualizing a Document Collection. *2012 45th Hawaii International Conference on System Sciences*, 1835–1844. doi:10.1109/HICSS.2012.57
- Gorg, C., Liu, Z., Parekh, N., Singhal, K., & Stasko, J. (2007). Visual analytics with jigsaw. *2007 IEEE Symposium on Visual Analytics Science and Technology* (pp. 201–202). IEEE. doi:10.1109/VAST.2007.4389017

- Gotz, D. (2007). The ScratchPad: sensemaking support for the web. *Proc. of the Inter. WWW Conf. Posters*, (1), 1329–1330. Retrieved from http://smtp.gotzfamily.org/~gotz/research/publications/gotz_www_2007.pdf
- Gotz, D., Zhou, M., & Aggarwal, V. (2006). Interactive Visual Synthesis of Analytic Knowledge. *2006 IEEE Symposium On Visual Analytics And Technology* (pp. 51–58). IEEE. doi:10.1109/VAST.2006.261430
- Heer, J., Viégas, F. B., & Wattenberg, M. (2007). Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (p. 1029). New York, New York, USA: ACM Press. doi:10.1145/1240624.1240781
- Heuer Richards J., J. (1999). *Psychology of Intelligence Analysis*. Government Printing Office.
- IEEE VAST Challenge 2011. (2011). Retrieved September 15, 2012, from <http://hcil.cs.umd.edu/localphp/hcil/vast11/>
- Isenberg, P., Tang, A., & Carpendale, S. (2008). An exploratory study of visual information analysis (pp. 1217–1226). Florence, Italy: ACM. Retrieved from <http://portal.acm.org.proxy.lib.sfu.ca/citation.cfm?id=1357054.1357245&coll=ACM&dl=ACM&CFID=66675556&CFTOKEN=23444201>
- Jonker, D., Wright, W., Schroh, D., Proulx, P., & Cort, B. (2005). Information Triage with TRIST. *2005 International Conference on Intelligence Analysis*. Retrieved from http://vadl.cc.gatech.edu/documents/54_Wright_Oculus_TRIST_Final_Distrib.pdf
- Kadivar, N., Chen, V., Dunsmuir, D., Lee, E., Qian, C., Dill, J., Shaw, C., et al. (2009). Capturing and supporting the analysis process (pp. 131–138).
- Liu, Z., Gorg, C., Kihm, J., Lee, H., Choo, J., Park, H., & Stasko, J. (2010). Data ingestion and evidence marshalling in Jigsaw VAST 2010 Mini Challenge 1 award: Good support for data ingest. *2010 IEEE Symposium on Visual Analytics Science and Technology* (pp. 271–272). IEEE. doi:10.1109/VAST.2010.5653042

- Mahyar, N., Sarvghad, A., & Tory, M. (2010). A closer look at note taking in the co-located collaborative visual analytics process. *IEEE VAST*, 171–178. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5652879
- Marshall, C. (1990). WORK PRACTICES STUDY : ANALYSTS AND NOTETAKING. *Unpublished report*, (May). Retrieved from <http://www.csd.tamu.edu/~marshall/oswrstudy.pdf>
- Nobarany, S., Haraty, M., & Fisher, B. (2012). Facilitating the reuse process in distributed collaboration. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12* (pp. 1223–1232). New York, New York, USA: ACM Press. doi:10.1145/2145204.2145388
- Pacific National Northwest Laboratory. (2013). IN-SPIRE™. Retrieved from <http://inspire.pnnl.gov/>
- Pirolli, P., & Card, S. K. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Analyst, The*. Retrieved from https://analysis.mitre.org/proceedings/Final_Papers_Files/206_Camera_Ready_Paper.pdf
- Quinlan, P. R., Reed, C., & Thompson, A. (2008). INSPIRE: An Integrated Agent Based System for Hypothesis Generation within Cancer Datasets (Vol. 3, pp. 587–590).
- Redish, J. (2007). Expanding usability testing to evaluate complex systems. *Journal of Usability Studies*, 2(3), 102–111. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.135.129&rep=rep1&type=pdf>
- Robinson, A. C. A. (2008). Collaborative synthesis of visual analytic results. *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on* (pp. 67–74). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4677358

- Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools: Multi-dimensional In-depth Long-term Case. *Proceedings of the 2006 AVI workshop on BEyond time and errors novel evaluation methods for information visualization - BELIV '06*, 1. doi:10.1145/1168149.1168158
- Shrinivasan, Y. B., & Van Wijk, J. J. (2008). Supporting the analytical reasoning process in information visualization. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08* (pp. 1237–1246). New York, New York, USA: ACM Press. doi:10.1145/1357054.1357247
- Stasko, J., Görg, C., & Spence, R. (2008). Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2), 118–132. doi:10.1057/palgrave.ivs.9500180
- Tableau Software. (2011). Tableau Desktop. Retrieved from <http://www.tableausoftware.com/>
- Thomas, J. J., & Cook, K. A. (2005). *Illuminating the Path: The R&D Agenda for Visual Analytics: National Visualization and Analytics Center (NVAC)*. Retrieved from <http://nvac.pnl.gov/agenda.stm>
- Willett, W., Heer, J., Hellerstein, J., & Agrawala, M. (2011). CommentSpace: structured support for collaborative visual analysis. *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11* (pp. 3131–3140). New York, New York, USA: ACM Press. doi:10.1145/1978942.1979407
- Wong, P. C., Hetzler, B., Posse, C., Whiting, M., Havre, S., Cramer, N., Shah, A., et al. (2004). IN-SPIRE InfoVis 2004 Contest Entry, 216.2. doi:10.1109/INFOVIS.2004.37
- Wright, W., Schroh, D., Proulx, P., Skaburskis, A., & Cort, B. (2006). The Sandbox for analysis: concepts and methods. *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 801–810). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1124890>

Appendices

Appendix A. Log File Tags

The following table lists the additional tags that are outputted in the log file and a description of what triggers the tag output.

Creating New Notes

Tags	Trigger for the Tag Output
[create-note] Created_In[hybrid-view] Using_[click-in-graph]	A blank note is created in the <i>Hybrid Graph View</i>
[create-note] Created_In[hybrid-view][note-repo] Using_[create-note-dialog]	A blank note is created in the notes repository of the <i>Hybrid Graph View</i>
[create-note] Created_In[semantic-zoom-view][note-repo] Using_[create-note-dialog]	A blank note is created in the notes repository of the <i>Semantic Zoom View</i>
[create-note] Created_In[semantic-zoom-view] Using_[create-note-dialog]	A blank note is created in the <i>Semantic Zoom View</i>
[attach-note] Attach_In_View_[hybrid-view] Attach_To[graph-nodes]	A note is attached to the graph nodes in the <i>Hybrid Graph View</i>
[attach-note] Attach_In_View_[hybrid-view] Attach_To[graph-nodes] [note-reference-note]	A note is attached to another note in the <i>Hybrid Graph View</i>
[attach-note] Attach_In_View_[document-view] Attach_To[document] [attach-to-text]	A note is attached to the text of a document in the <i>Document View</i>
[append] [attach-note] Attach_In_View_[document-view] Attach_To[document] [note-reference-note] [attach-to-text]	A note is attached to the text of a another note in the <i>Document View</i>
[attach-note] Attach_In_View_[hybrid-view] [note-repo] Attach_To[note-panel] [note-reference-note] [attach-to-text]	A note is attached to the text of another note in the Notes Repository in the <i>Hybrid Graph View</i>
[append] [attach-note] Attach_In_View_[semantic-zoom-view]	A note is attached to the text of another note in the Notes Repository in the

[note-repo] Attach_To[note-panel] [note-reference-note] [attach-to-text]	<i>Semantic Zoom View</i>
[attach-note] Attach_In_View_[hybrid-view] Attach_To[note-panel] [note-reference-note] [attach-to-text]	A note is attached to the text of another note in the <i>Hybrid Graph View</i>
[append] [attach-note] Attach_In_View_[semantic-zoom-view] Attach_To[zoom-document] [note-reference-note] [attach-to-text]	A note is attached to the text of another note in the <i>Semantic Zoom View</i>

Adding references to a note

Tags	Trigger for the Tag Output
[add-reference-in-note] Add_In_View_[document-view] Using_[extract-entity-from-non-note]	Entity extracted from a document in the <i>Document View</i>
[add-reference-in-note] Adding_In_[hybrid-view] Using_[note-entity-suggestion]	An entity was added as a reference using the entity suggestion feature in a <i>CZNote</i> panel

Attaching a note to a note

Tags	Trigger for the Tag Output
[add-reference-in-note] Adding_In_[hybrid-view] Using_[note-add-link]	An entity/document was added as a reference to a note using the add link icon in the <i>Hybrid Graph View</i>
[attach-note] Attach_In_View_[hybrid-view] [note-repo] Attach_To[note-panel] [note-reference-note] [attach-to-text]	A note was attached to the text of another note in a notes repository

How were notes added to a data view?

Tags	Trigger for the Tag Output
[add-to-view] [notes-only] To_[hybrid-view] From_[hybrid-view] [note-repo] Using_[note-repo][drag-and-drop]	Notes were added to a <i>Hybrid Graph View</i> from a notes repository
[add-to-view] [notes-only] To_[semantic-zoom-view] From_[semantic-zoom-view] [note-repo] Using_[note-repo][drag-and-drop]	Notes were added to a <i>Semantic Zoom View</i> from a notes repository
[add-to-view] [notes-only] To_[document-view] From_[semantic-zoom-view] [note-repo] Using_[note-repo][drag-and-drop] [add-to-view] [non-notes] To_[document-view] From_[hybrid-view] Using_[drag-and-drop] [add-to-view] [notes-mixed] To_[document-view] From_[hybrid-view] Using_[drag-and-drop]	Notes non-notes notes-mixed were added to a <i>Document View</i> from a <i>Hybrid Graph View</i> using drag and drop
[add-to-view] [notes-only] To_[hybrid-view] From_[hybrid-view] Using_[drag-and-drop] [add-to-view] [non-notes] To_[document-view] From_[hybrid-view] Using_[drag-and-drop] [add-to-view] [notes-mixed] To_[hybrid-view] From_[hybrid-view] Using_[drag-and-drop]	Notes non-notes notes-mixed were added to a <i>Hybrid Graph View</i> using drag and drop

<p>[add-to-view] [notes-only] To_[semantic-zoom-view] From_[hybrid-view] Using_[drag-and-drop]</p> <p>[add-to-view] [non-notes] To_[semantic-zoom-view] From_[hybrid-view] Using_[drag-and-drop]</p> <p>[add-to-view] [notes-mixed] To_[semantic-zoom-view] From_[hybrid-view] Using_[drag-and-drop]</p>	<p>Notes non-notes notes-mixed were added to a <i>Semantic Zoom View</i> using drag and drop</p>
<p>[add-to-view] [non-notes] To_[CZDocumentView3] From_[search-panel] Using_[search-panel]</p> <p>[add-to-view] [notes-mixed] To_[CZHybridView2] [CZSemanticZoomView1] From_[search-panel] Using_[search-panel]</p> <p>[add-to-view] [notes-only] To_[SemanticZoomView3] From_[search-panel] Using_[search-panel]</p>	<p>Notes were added to the <i>Document View</i> <i>Hybrid Graph View</i> <i>Semantic Zoom View</i> from the search panel</p>
<p>[add-to-view] [notes-only] To_[document-view] From_[hybrid-view] Using_[entities-popup-menu]</p> <p>[add-to-view] [notes-mixed] To_[hybrid-view] From_[hybrid-view] Using_[entities-popup-menu]</p> <p>[add-to-view] [non-notes] To_[semantic-zoom-view] From_[hybrid-view] Using_[entities-popup-menu]</p>	<p>Notes were added to the <i>Document View</i> <i>Hybrid Graph View</i> <i>Semantic Zoom View</i> using the entities popup menu</p>
<p>[add-all-entities-by-type] [non-notes] Added_In[hybrid-view]</p> <p>[add-all-entities-by-type] [notes-mixed] Added_In[hybrid-view]</p>	<p>All note entities non-note entities notes-mixed were added to the <i>Hybrid Graph View</i></p>

[add-all-entities-by-type] [notes-only] Added_In[hybrid-view]	
[add-all-entities-by-type] [notes-mixed] Added_In[semantic-zoom-view] [add-all-entities-by-type] [notes-only] Added_In[semantic-zoom-view] [add-all-entities-by-type] [non-notes] Added_In[semantic-zoom-view]	All note entities non-note entities notes-mixed were added to the <i>Semantic Zoom View</i>
[add-all-entities-by-type] [notes-mixed] Added_In[document-view] [add-all-entities-by-type] [notes-only] Added_In[document-view] [add-all-entities-by-type] [non-notes] Added_In[document-view]	All note entities non-note entities notes-mixed were added to the <i>Document View</i>

Searching

Tags	Trigger for the Tag Output
[search-result] Search_For_[Nicolai] Search_From_view.notes.repositorypanel. CZNoteRepo[,2,25,262x560,invalid,layout=j ava.awt.GridBagLayout,alignmentX=0.0,ali gnmentY=0.0,border=,flags=9,maximumSiz e=,minimumSize=,preferredSize=]	Searches were made in the notes repository
[search-result] Search_For_[Boonmee] Search_From_[search-panel] With_Results_[non-notes] [search-result] Search_For_[Nicolai] Search_From_[search-panel] With_Results_[notes-only] [search-result] Search_For_[Nicolai]	A search returned results with non-notes only notes-only mix of notes and non-notes

<p>Search_From_[search-panel] With_Results_[notes-mixed]</p>	
<p>[prepend] Using_[double-click-on-nodes] [hybrid-view] [get-related-from-non-note-entities] With_Results_[non-notes] [prepend] Using_[double-click-on-nodes] [hybrid-view] [get-related-from-non-note-entities] With_Results_[notes-mixed] [prepend] Using_[double-click-on-nodes] [hybrid-view] [get-related-from-non-note-entities] With_Results_[notes-only] [prepend] Using_[add-node-panel] [hybrid-view] [get-related-from-notes-and-mixed-entities] With_Results_[notes-only] [prepend] Using_[double-click-on-nodes] [hybrid-view] [get-related-from-notes-and-mixed-entities] With_Results_[notes-mixed] [prepend] Using_[double-click-on-nodes] [hybrid-view] [get-related-from-notes-and-mixed-entities] With_Results_[non-notes] [prepend] Using_[double-click-on-nodes] [hybrid-view] [get-related-from-notes-only] With_Results_[non-notes] [prepend] Using_[double-click-on-nodes] [hybrid-view] [get-related-from-notes-only] With_Results_[notes-mixed] [prepend] Using_[double-click-on-nodes] [hybrid-view] [get-related-from-notes-only] With_Results_[notes-only]</p>	<p>Get related entities from non-note-entities notes-only mixed notes and non-note entities</p>

[hybrid-view] [get-edges-with-notes-and-mixed-entities]	Find the edges between mixed note and non-note entities notes-only non-note entities
[hybrid-view] [get-edges-with-notes-only]	
[hybrid-view] [get-edges-with-non-note-entities]	

Entity refinements

Tags	Trigger for the Tag Output
[merge] [document] [entities-list-panel] [entities-popup-menu]	Merging entities within a <i>Document View</i>
[merge] [hybrid-view] [graph-nodes]	Merging entities within a <i>Hybrid Graph View</i>
[append] Remove_From_[non-note] [remove-entity-single] [append] Remove_From_[non-note] [remove-entity-all] [append] Remove_From_[note] [remove-entity-single] [append] Remove_From_[note] [remove-entity-all]	Removing an entity from a note non-note

Appendix B. Study Description

How does the integration of data and notes affect the analysis process in a visual analytic tool, CZSaw?

Principal Investigators: Eric Lee ([eric.lee@umd.edu](#)), John Dill ([john.dill@umd.edu](#))

Co-Investigators: David Botta ([david.botta@umd.edu](#)), Robert Woodbury ([rob.woodbury@umd.edu](#))

Study Description

As a member of an analysis team, you will use CZSaw and CZNotes to help solve the IEEE VAST 2011 Mini Challenge 3 (MC3) over the course of 3 weeks. The official VAST 2011 MC3 task description can be found here: <http://hcil.cs.umd.edu/localphp/hcil/vast11/index.php/taskdesc/index>

1. You will use CZSaw and CZNotes to analyze the challenge data set individually and collaboratively with other participants as a team.
2. You will fill out a journal following any individual and collaborative analysis sessions. The journaling is open-ended but there are also guiding questions to record your activities performed during your analysis sessions.
3. You will also be asked to participate in 1-2 short group debrief sessions per week to share analysis progress and feedback on CZNotes usage. The investigators will be present to facilitate these sessions. These sessions will be video recorded.
4. Data collection will be in the form of video and audio recording during the group debrief sessions, journaling, and logs of your interactions with CZSaw and CZNotes recorded by the CZSaw software.
 - a. Please bring your project files on the last debrief session of each week OR upload to server (to be determined later)
5. The final product of this analysis task is to provide a "Detailed Answer", as a group, of your analysis of the VAST 2011 Mini Challenge 3 data as per the official instructions (found in the above task description link).
 - a. "A Detailed Answer is a longer text description focusing on how you arrived at the answer with much more details than the Short Answer. For mini challenges, detailed answers are limited to 1000 words (including captions), with a maximum of 5 screen shots. Detailed answers should provide the answer and describe in detail the PROCESS USED TO ARRIVE AT THE ANSWER."

Questionnaires

Pre-study Questionnaire

There is a pre-study questionnaire that you will complete online at:

<http://websurvey.sfu.ca/survey/86915474>

Post-study Questionnaire

A link to an online post study questionnaire will be provided to you at a later date.

How will the analysis process be affected by representing analytical findings (notes) as entities in an entity based visual analytic tool?

1 of 3

Journaling

We would like you to journal your activities after each analysis session. You can journal your analysis activities using either one of these methods or a combination of:

1. The provided journaling book
2. An online websurvey: <http://websurvey.sfu.ca/survey/87857911>
3. Your own method of journaling

Please note that your journals must be in a collectable form, as they will be collected by the Investigator for data analysis at the end of this study.

Please include the following topics in your journaling. Give brief description of what you did in the analysis session; include such things as

1. What did you do in this analysis session? What was your goal and what was accomplished?
2. Collaboration (who, what, when, where, why, how?)
3. How did you use CZNotes? (eg to bring in new data, identify relationships)
4. Did you perform any entity refinements? where and what type? (e.g. merging entities in a document, extracting a new entity in a note)
5. Do you have any other comments? or feedback on CZNotes?

Grouping Sessions

In the grouping sessions, you can spend time to discuss your analysis progress as a group. We will also spend time to discuss any comments or problems with using CZNotes and CZSaw during your analyses. These sessions are video and audio recorded.

Collecting Your CZSaw Project Folders

Please do not delete any of your project folders. They will be collected on a weekly basis by the investigator.

How to use CZNotes

Please refer to the CZNotes tutorial videos here:

<http://www.youtube.com/watch?v=uW9af39wGs0>

<http://www.youtube.com/watch?v=T3wzHWaqR1w>

If you have any questions, please do not hesitate to contact the investigator at ela10@sfu.ca.

If you encounter any problems or bugs with CZNotes or CZSaw, you can also report the bug through our user feedback tool detailed below.

Bug Reporting and Bug Fixes

Redmine will be used for bug reporting and tracking. You may post a new issue here:

<http://hvilab.iat.sfu.ca/redmine/projects/feedback/issues/new>

Please indicate the CZSaw build number when posting an issue. Also, upload a zip of your project folder. If the project folder is too large, please upload the script.txt and log.txt files from your project folder.

New builds will be available when bug issues are fixed. You will be notified by email when a new build is available for use.

Script Backups

Your project scripts are automatically backed up as text files with a time-stamped file name in a folder called "script_backup" in the project folder. Scripts are backed up after every 5 transactions are executed in CZSaw as well as when you open and close a project. In the case of damaged or lost script files, you can copy a backup script file from the "script_backup" folder into the root directory (with the backup script renamed to "script.txt").

Exporting/Importing Data Commands from your Script

All data commands such as extracting a new entity, merging entities as well as creating and editing a note can be exported in a "barebones" script containing only these script commands (under the Script menu in CZSaw). Thus, you can create a new project and import such a script to keep all your data editing commands as well as import all the notes you have created from another project into a new project.

Appendix C. Pre-Study Questionnaire

Q1. Enter your participant ID

Answer :

Q2. Please fill in the following background information

What is your age? :

Q3. What is your gender?

Male

Female

Q4. Please provide us your educational background information

Institutions :

Programs :

Current Year :

Q5. What is your level of experience with using Visual Analytic software?

Q6. What is your level of experience with using CZSaw?

Q7. Did you use notes of any kind in previous problems? What sort of notes and how did you make them?

Q8. How did you use the notes you made?

Q9. Did you share your notes with others? If so, how?

Appendix D. VA Tool Familiarity Questionnaire

Q1. What is your participant ID

Answer :

Q2. How knowledgeable are you about: (None, Some, Moderate, Knowledgeable)

The Analysis Process? (e.g. Pirolli and Card's Sensemaking Model) :

Analysis Techniques? (e.g. Heuer's Analysis of Competing Hypotheses) :

Q3. How experienced are you with the following Visual Analytic tools?

(None, A bit, Moderate, Expert)

Tableau :

IN-SPIRE :

Geo-Time :

Jigsaw :

CZSaw :

Starlight :

Appendix E. Post-Study Questionnaire

Q1. What is your participant ID?

Answer :

Q2. What features of *CZNotes* supported your analysis tasks and how?

Q3. What did you not like about *CZNotes* (e.g. its functionalities, its properties, user interface, etc.)? Were there any additional features you would have liked *CZNotes* to have?

Q4. Did you find it beneficial for notes to be represented the same way as the data (i.e. *CZNotes* as another entity type)?

Not Useful

Somewhat Not Useful

Indifferent

Somewhat Useful

Very Useful

Q5. Please elaborate on the usefulness/non-usefulness of notes being represented the same way as the data in *CZSaw* (i.e. *CZNotes* as another entity type)

Q6. Did you use the data views (*Semantic Zoom View*, *Hybrid Graph View*, *Document View*) to visualize and interact with your notes? If so, how and why?

Q7. Please provide any other observations/opinions/feedback you feel would be of help with regard to further development of the Notes capability.

Appendix F. Participant Consent Form

How does the integration of data and notes affect the analysis process in a visual analytic tool, CZSaw?

Principal Investigators: Eric Lee (), John Dill ()
Co-Investigators: David Botta, Robert Woodbury

The purpose of this study:

The study investigates how integrating note-taking capabilities, CZNotes, in an existing visual analytics tool, CZSaw, impacts the analysis of unstructured text data. CZSaw is a computer software tool used to support the analysis of text data and CZNotes is a component of CZSaw that allows users to capture notes within the tool. We will observe how a group of users use CZNotes individually and collaboratively to solve the IEEE VAST 2011 Mini Challenge 3, which involves analyzing a fictional set of news reports. In particular, we are interested in how CZNotes impacts the process of analysis.

Your participation includes:

As a member of an analysis team, you will use CZSaw and CZNotes to help solve the IEEE VAST 2011 mini Challenge 3 over the course of 3 weeks. You will use CZSaw and CZNotes to analyze the challenge data set individually and collaboratively with other participants. You will fill out a questionnaire following your individual and collaborative analysis sessions. The questionnaire pertains to recording when and how you performed your analysis sessions. You will also be asked to participate in 2-3 short group debrief sessions per week to share analysis progress and feedback on CZNotes usage. The investigators will be present to facilitate these sessions. You will be informed about how and when data will be collected from you. Data collection will be in the form of video and audio recording during debrief sessions, periodic questionnaires, and logs of your interactions recorded by the CZSaw software.

Free to refuse or withdraw:

You can withdraw from this study at any time after agreeing to participate. If you refuse to continue, or withdraw, there will be no consequences. However, the compensation offered will be prorated based on the time that you participated. Your withdrawal will not have any negative impact on your academic grades.

Your privacy and confidentiality:

Data collected during this study will be secured in a locked room and locked cabinet and on password-secured storage devices using TrueCrypt. Your data will remain anonymous in that the data collected from you will be associated with you through a random participant ID number. Your identity will never be revealed in publications or presentations. You will be asked for specific permission if the investigators wish to use any audio or video of you from the debrief sessions in publications or presentations. The collected data will be kept for the duration of the research project and will be deleted and shredded after two years (December 2013).

Risks and benefits:

There are no expected risks nor benefits to you for participating in this study. Your participation may contribute knowledge to the field of visual analytics.

Appendix G. Sample PERL Script used in Data Analysis

```
# Perl trim function to remove whitespace from the start and end of the string
sub trim($)
{
    my $string = shift;
    $string =~ s/^\s+//;
    $string =~ s/\s+$//;
    return $string;
}

our @dName;

my @dirs = grep { -d } glob "/Thesis/User-Study/Collected-Data/Logs/Logs-No-Replay/**";

foreach (@dirs) {
    #print $_."<curr\n";
    @dName = ($_ =~ m/(P\d)/g);
    #print "@dName\n";

    #loop through each log file in a directory
    opendir(DIR, $_) or die;
    while (defined(my $file = readdir(DIR))) {
        # do something with "$dirname/$file"
        next unless $file =~ /\.txt$/;
        #print "Transactions from $file\n";

        #combine all the lines of the file into a single variable, $content
        open LOGFILE, "$_/$file" or die "can't open $file";
        my @lines = <LOGFILE>;
        my $content = join(" ", @lines);

        # my @timestamps = ($content =~ m/[(.*)]/g);
        # print "@timestamps\n";
        #print @dName."<innerdName\n";
        #setup output file for the results
        open (OUTPUTFILE, ">>tagsOutput"."@dName".".txt");

        while((my $key, my $value) = each(%regExList)) {
            #find all the timestamps (RegeX group 1) and its corresponding transaction name (RegEx
            group 2)

            while ($content =~ m/($key)/g)
```

```

        {

            my $tag = $1;

            $tag =~ m/^[(.*)\[\s](.*)/g;

            my $timestamp = trim($1);
            my $match = trim($2);

            #separate the matching groups by a pipe
            print OUTPUTFILE "$value|$file|$timestamp|$match|@dName|\n";

        }
    }

    close LOGFILE;

    #for each timestamped transaction

    #-time and transaction name as a pair

    #capture new project and close project tag and calculate duration

    #capture open project and close project tag and calculate duration

    #output all into a single spreadsheet or csv file

}

close OUTPUTFILE;

#process the output file from above
open OUTPUT, "tagsOutput"."@dName". ".txt" or die;
my @lines = <OUTPUT>;
my $content = join(" ", @lines);

#setup output file for the results
#open (OUTPUTFILE, ">>CSVtagsOutput"."@dName". ".csv");
open (OUTPUTFILE, ">>CSVTagsOutput.csv");

if(! -s OUTPUTFILE) {
    print OUTPUTFILE "Tag|Log File Name|Timestamp|Match|Participant\n";
}

```

```
#separate each timestamp and transaction name pair into its own rows
while ($content =~ m/([^\|]*\|[^|]*\|[^|]*\|[^|]*\|[^|]*)/g)
{
    print OUTPUTFILE $1;
}

print OUTPUTFILE "\n";
close OUTPUT;

closedir(DIR);
}
```