Jigsaw: supporting investigative analysis through interactive visualization

John Stasko[1]
Carsten Görg[1]
Zhicheng Liu[1]

[1]*School of Interactive Computing & GVU Center, Georgia Institute of Technology, Atlanta, GA, U.S.A.*

Correspondence:
John Stasko, School of Interactive Computing & GVU Center, Georgia Institute of Technology, 85 5th St., NW, Atlanta GA 30332-0760, U.S.A. Tel.: +1 404 894 5617; Fax: +1 404 894 3146; E-mail: stasko@cc.gatech.edu

Abstract

Investigative analysts who work with collections of text documents connect embedded threads of evidence in order to formulate hypotheses about plans and activities of potential interest. As the number of documents and the corresponding number of concepts and entities within the documents grow larger, sense-making processes become more and more difficult for the analysts. We have developed a visual analytic system called `Jigsaw` that represents documents and their entities visually in order to help analysts examine them more efficiently and develop theories about potential actions more quickly. `Jigsaw` provides multiple coordinated views of document entities with a special emphasis on visually illustrating connections between entities across the different documents.
Information Visualization (2008) **7,** 118–132. doi:10.1057/palgrave.ivs.9500180

Introduction

Investigative analysts seek to make discoveries and uncover hidden truths from large collections of data and information. Scientists follow this process when they read research papers to learn about related efforts; news reporters perform such analyses when they investigate new stories; law enforcement and intelligence analysts carry out these kinds of investigations when they review case reports. One common element of all these analytic activities is that they are cognitively very challenging, frequently involving large collections of data and text that tax a person's memory, deduction, reasoning, and general analytic capabilities.

In the latter example above, law enforcement and intelligence, an investigative analyst gathers individual chunks of evidence that may range from incident reports filed by operatives in the field to open source news reports such as articles gathered via web searches. Often, the investigative process involves analysts pouring over large collections of evidence, reading and reviewing the documents to make connections between seemingly disparate facts, essentially 'reading between the lines' to link separate activities into a larger plot or narrative. Sometimes the connections can be clear, for instance, a particular individual mentioned in different reports. Alternatively, the connections can be more difficult to discern, for instance, approximate overlaps in time or location.

While reading documents and digesting the information therein, analysts gradually form internal mental models of the people, places, and events discussed in the documents. As the number of documents grows larger, however, it becomes increasingly difficult for an investigator to track the connections between data and make sense of it all. The sheer number of entities involved may make it very difficult for a person to form a

clear understanding of the underlying concepts and relationships in the document collection.

Investigative analysis activities are a major aspect of the broader field of sense-making.[1] Sense-making has been defined as 'A motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively'.[2] Sense-making activities certainly include the types of investigations on which we focus, but they include other activities as well, for instance, deciding which digital camera to buy or which hotel in London to stay at during a vacation.

At the core of our research is a belief, shared by many others, that visual representations of data and documents can help people better examine, analyze, and understand them. For example, Norman[3] has described how visual representations can help augment people's thinking and analysis processes. Card *et al*.[4] refer to visuals used in this manner as 'external cognition aids.'

The objective of our research is to develop visual representations of the information within textual document and report collections in order to help analysts search, review, and understand the documents better. We seek to create interactive visualizations that will highlight and identify connections between entities in the documents where entities may be people, places, dates and organizations, for instance. Fundamentally, we want to build visual representations of the documents that help analysts browse and explore them, making sense of all the facts and information contained therein.

Our goal is not to replace the documents, however. We firmly believe that analysts must carefully read documents to best understand them. What we seek to provide is a type of interactive visual index onto the documents, a visual analytic system[5,6] that connects and links entities discussed therein and thus guides analysts toward the documents to read next. Furthermore, the interactive visualizations should provide representations that assist analysts in building accurate and informative conceptual models of the underlying themes, plots, and stories embedded in the document collection. In our experience, investigative analysts have expressed a reluctance to use 'black-box' systems; they instead prefer a more transparent explorative method. Our approach is decidedly human-centered but is augmented with analytic support; we want to design an easy-to-use system that puts the analyst in charge.

Pirolli and Card[7] performed a cognitive task analysis of intelligence analysts and their work that resulted in a notional model of the intelligence analysis process. Their model is organized around two major activity loops, foraging and sense-making. Our work touches on both loops, helping analysts to choose useful documents to examine next and also to develop schema and hypotheses that fit the available evidence. Pirolli and Card identify several leverage/pain points particularly in need of assistance within analytic processes. Two, in particular, that our work addresses involve (1) the cost structure of

scanning and selecting items for further attention and (2) analysts' span of attention for evidence and hypotheses. They comment on the two leverage points, respectively:

'Our analysts spent considerable time scanning data seeking relevant entities (names, numbers, locations, etc.). The assessment of whether or not an item is relevant also takes time. Techniques for highlighting important information with pre-attentive codings, or re-representing documents (e.g., by summaries) appropriate to the task can improve these costs.'

'Techniques aimed at expanding the working memory capacity of analysts by offloading information patterns onto external memory (e.g., visual displays) may ameliorate these problems.'

To address such objectives, we have designed a suite of interactive visualizations and built a prototype system called `Jigsaw` that implements the visualizations as separate views onto a (text) document collection. The views are connected so that actions within one view can be reflected in the others. We named the system `Jigsaw` because we think of all the different entities and facts in a document collection as the pieces of a puzzle. The `Jigsaw` system should help an analyst 'put the pieces together.'

Our research adopts Pirolli and Card's[7] conceptual model of intelligence analysis. We also draw upon the work of Heuer[8] and Johnston[9] to help understand the processes and practices of analysts. Furthermore, our own earlier research on task analysis for information visualization[10,11] and on the importance and use of interaction in information visualization[12] directly influences the design of `Jigsaw`.

This article is an expanded and updated version of a paper[13] presented at the VAST 2007 Symposium. In the next section, we provide more details about the types of documents that are the focus of analysis in `Jigsaw`. We also describe the importance of identifying the different types of entities in a document since these entities serve as the primary basis for the visualizations. The System Description section reviews the `Jigsaw` system in detail, its underlying data structures, system architecture, event messaging, and each of the different views. In the subsequent section we provide a short example scenario of use to better help the reader understand how the system functions. A video that demonstrates the scenario actions is available on the `Jigsaw` website.[14] In the VAST 2007 Contest section we describe our experience about participating in the VAST 2007 Contest. In the next section we outline other, more general sense-making domains, `Jigsaw` could be applied in. The paper concludes with a discussion of related work and a list of ongoing and future efforts planned for the system.

Analyzing documents

The target artifact of our study is a textual document describing a set of facts or observations from the domain

of interest. We assume that the documents will be in a natural language, loose narrative format, and will be about 1–6 paragraphs. While our visualization techniques still operate for larger documents, pragmatically the techniques become less useful as the document size increases because the higher number of entities per document swamps the display. Our intended target is a smaller document with a few nuggets of information contained therein. News stories and case reports are good examples of the type of document we focus on, and this type of document is often the focus of analysts in law enforcement or intelligence.[5,15]

Analysis can draw on data from varied and distributed sources. The distributed nature of information leads to heterogeneity across the documents in terms of topic, authors, content, style, date, and so on. Furthermore, different reports will contain information that is unclear, confusing, or even contradictory. Organizational tools have to consider both the complexity of the information and the analysis task.

Below is an example document, taken from the VAST 2007 Conference Contest,[16] that provides a flavor of the types of documents on which we are focusing. A large number of different events, items, themes, and stories can be embedded throughout a collection of thousands or even just hundreds of such reports.

Report 20040510-4_16: May 14, 2004

```
VANCOUVER, British Columbia -- A Canadian
immigration panel is considering whether
accused environmental saboteur Tre Arrow can
apply for refugee status in Canada.

Arrow, 30, who is wanted for fire bombing
logging and cement trucks in Oregon, asked the
Canadian authorities to remain in Canada as a
political refugee at a hearing in Vancouver on
Tuesday.

A key issue will be whether Arrow is affiliated
with a terrorist group, which would immediately
disqualify him from receiving refugee status in
Canada, authorities said.

The Immigration and Refugee Board is scheduled
to decide by May 31 whether Arrow is affiliated
with the Earth Liberation Front, a group the
FBI considers a terrorist organization respon-
sible for scores of attacks on property over
the past dozen years.
```

Each document in the collection includes a number of facts and refers to people, places, objects, dates, actions, and so on. Individual documents are relatively focused, however, and typically describe a particular event or occurrence. So, one may think of each document as presenting a small bit of evidence. Thus, investigative analysis in our context is the process of connecting a series of individual bits of evidence to construct a larger, broader story or narrative. It is about understanding how the individual events and entities referred to in the documents relate to each other and, when composed together, reveal a larger plot.

For instance, consider police investigators who have a large collection of case reports. Suppose that one report describes a crime in which a blue Ford car was seen fleeing from the scene. Investigators may wonder if other case reports mention a blue Ford, and if so, what related activities are mentioned in those reports. Suppose that another such report exists and it describes a known criminal who used a blue Ford in a recent theft. Investigators might extend their search to examine activities of this individual, places he has been, known accomplices, and so on, keeping a careful watch for unexpected, serendipitous connections and coincidences that arise (and are not initially uncovered by a simple 'blue Ford' search). Ultimately, investigators are likely seeking to uncover larger criminal plans and threats in order to either thwart planned crimes before they occur or to apprehend the criminals being sought.

While other well-known systems such as IN-SPIRE[17] focus mainly on themes or concepts across document collections, the primary unit of analysis from documents for Jigsaw is an entity. Within any document one can identify a set of entities. Example entity types include person, place, date, organization, time, and money. Jigsaw allows any type of entity to be included in the analysis.

The goal of the Jigsaw system is to highlight and communicate connections and relationships between entities across a document collection. We believe that these connections, when assimilated, help to provide the analyst with a better global understanding of the broader themes and plans hinted at by the particular events and facts documented in the reports.

Obviously, an initial requirement for Jigsaw is to identify and extract the entities from each document and, ideally, store them in a format that allows easier analysis and manipulation. Named entity identification, however, is not the focus of our work, but it is a process that has been studied extensively.[18,19] We presently incorporate software from the ANNIE package of the GATE system[20] into Jigsaw to perform entity identification and we allow the analyst to manually identify or correct the identification of entities too. Alternative free and commercial entity identification systems also exist, such as Balie,[21] FreeLing,[22] Connexor,[23] and Cicero.[24] We chose GATE because it is open source and built in the Java programming language as is Jigsaw.

Further automated analysis of the document text could be performed as well. For instance, algorithms determining the content of documents with respect to frame analysis and social movement theory exist,[25] which tells us how people try to influence target audiences and how those audiences respond. Other computational linguistics analysis techniques exist too, but presently Jigsaw does

not incorporate this type of analysis. We believe that this is a fruitful area for future research.

We adopted analysis exercises created by Frank Hughes of the Joint Military Intelligence College[26] to serve as initial sample test data for `Jigsaw`. The exercises involve collections of fabricated reports with an embedded master plot. Different reports in the collection hint at this plot and the goal of the exercise is to discover and articulate the plot. To make identifying the master plot more challenging, threads of other unrelated plots are suggested in the reports as well. Analysts in training perform exercises like this as part of their educational process.

The 2006 and 2007 IEEE VAST Symposium Contests provide large synthetic document collections of this type as well. We entered the 2007 Contest using `Jigsaw` as an analytic aid and won the university component of the contest. In fact, we use documents from the 2007 VAST Contest in the descriptions of `Jigsaw` and its views in the next section. The sample analytic scenario described later in the fourth section uses a report set from one of the Hughes' exercises.[26] We have altered some of the names and other entities from the already fabricated documents of that exercise for further anonymization.

System description

Overview

`Jigsaw` provides an analyst with multiple perspectives on a document collection. The system's primary focus is on displaying connections between entities across the documents and providing a type of visual index onto the document collection. We use a simple and easy-to-understand model of entity 'connection': two entities are connected if they appear in one or more documents together. Other, more semantically rich models of connection could be incorporated into `Jigsaw` as well.

`Jigsaw` presents information about documents and entities through multiple distinct visualizations, called views. Each view provides a different perspective onto the data. The views, which are discussed in more detail in the following subsections, include:

- a List View containing multiple reorderable lists of entities in which connections between entities are shown by coloring related entities and drawing links between them, multiple sorting options are available;
- a Graph View displaying connections between entities and documents in a node–link diagram, allowing analysts to dynamically explore the documents by showing and hiding links and nodes;
- a Scatter Plot View highlighting pairwise relationships between any two entity types and supporting focus on a specific subset of the displayed entities using range sliders;
- a Document View displaying the original text document with highlighted entities, showing how often a

document already has been viewed, and supporting entity modification;
- a Calendar View providing an overview of the documents and the entities within them according to the publication date of the document;
- a Document Cluster View representing all documents in the collection and providing manual and automated commands to partition the documents in meaningful clusters;
- a Shoebox supporting the analyst in the evidence marshalling process, providing functionality to build hypotheses and organize collected evidence.

The views implement two different perspectives on the document collection. The Scatter Plot View, the Document View, and the Document Cluster View are document based: documents are units of interaction and entities such as place or person are only shown in the context of a document. The List View, the Graph View, and the Calendar View, on the other hand, explicitly present entities as well as documents as units of interaction.

While designing the `Jigsaw` system we focused on keeping the user interface as easy to use as possible. The most common operations are available by single and double mouse clicks. A single click selects an item, a double click expands an item. More specifically, expanding a document shows all its contained entities and expanding an entity shows all the documents in which it can be found. Each of the views handles the two events differently and provides its own style of visual feedback. Other advanced operations, such as displaying items in other views or removing items from a view, are available in a right click-initiated pop-up menu.

User interaction with one view is translated to an event and communicated to all other views which then update themselves appropriately. Through such communication, different aspects of the documents can be examined simultaneously under different perspectives. For each view, users can turn on and off event receipt depending upon whether they want the view to stay synchronized with the most recent interactions in other views (the satellite dish icon in the right upper corner of each view displays its event receiving state). Users can create multiple instances of each view type and clone existing views as well. This capability allows a view to be frozen at an interesting state (event receipt turned off) while a new version of that view continues to receive events and update its state. Additionally, the user can bookmark a view which saves the current state of the view so that it can be returned to later in the investigative process. This operation provides a first step for asynchronous collaboration using `Jigsaw`: via bookmarks analysts can save and exchange views that document important states during an investigation.

Owing to the large amount of screen real estate required to display its views, `Jigsaw` ideally should be used on a computer with multiple and/or high-resolution monitors. We run the system on a computer with four

Figure 1 Jigsaw being used on a multiple-monitor computer.

displays, shown in Figure 1. Such a configuration allows one to minimize window flipping and reordering during analysis. Decreasing prices and smaller footprints of LCD monitors have made such configurations more common.

System architecture and infrastructure

Jigsaw is written in Java and adopts a model–view–controller architecture that separates the data (model) and user interface (view) components. Jigsaw creates objects for all documents and extracted entities and stores them in a general data structure model – the entities are tagged by their respective category name and bundled per document. This data structure is encapsulated in a class that provides an interface for the different view classes to call and retrieve entity-document information in order to build visualizations.

A controller class coordinates event communication from and to the views. Messages dispatched by views first go to the controller which then forwards the message to all receiving views. (Recall that each view provides a button to enable/disable event receiving.) A number of different event types exist: *select*, *display*, *expand*, *remove*, *entity-modification*, and *add-to-Shoebox*. A *select* event occurs when a user selects either an entity or a document in a view – such a selection is usually performed by a mouse click on the object. As feedback, the entity or document changes color or is highlighted visually. A *display* event occurs when a user explicitly indicates that an entity or a document should be displayed in all views where it is not currently visible. Users can initiate *display* events by performing a particular mouse gesture or by issuing a search query. An *expand* event is usually performed by a double click on an item and the views display the items connected to the one that was expanded. The remaining events occur when a user performs the corresponding command in the pop-up menu and the views update their state accordingly to the chosen command.

The Jigsaw Control Panel, shown in Figure 2(A), provides a variety of menu commands for use in the system. The active project (consisting of the document collection and the identified entities) and the workspace (consisting of all views that are currently open) can be saved and restored. New sources can be imported and added to the current or to a new project. Entities are extracted from the sources during the import process. Currently, Jigsaw can import plain text files or xml files with already identified entities. We plan to implement more filters to import webpages as well as pdf and Word documents.

The Control Panel also shows the color coding scheme for the available entity types and it provides a query interface for users to search for any entity as well as any string mentioned in the documents. When a query is issued, either the matching entities and/or the documents containing that string return as a result, and a display event is dispatched to the receiving views telling them to show the result(s). Finally, the Control Panel allows the user to instantiate the different views.

List View

The List View, illustrated in Figure 2(B), shows connections between sets of entities. The view consists of a number of lists, and each list shows entities of one specific type. The user can add and remove lists as desired – the number of lists is pragmatically constrained only by the horizontal space in the view. Once a list is displayed, a selection box at the top allows the user to change the entity type shown in that list. Thus, even the same type of entity can be placed side by side in the view; when investigating a social network, for example, it is useful to have two person lists side by side. The List View shown in Figure 2(B) contains three entity lists: places, persons, and organizations.

Selected entities are highlighted in bright yellow and all connected entities in all lists are highlighted in a shade of orange. The brightness of the highlighting on a connected entity indicates the strength of the connection: if the two appear together in only one document, a light orange is used, but if the two appear together in multiple documents, an increasingly dark orange is used as the number of co-appearances rises. Furthermore, the view draws lines between connected entities in adjacent lists to indicate the connection even further. In Figure 2(B), two entities are selected and highlighted in yellow: 'saboteur Tre Arrow' in the person list and the 'Immigration and Refugee Board' in the organization list.

The items in a list can be sorted either alphabetically, by the strength of the connection, or by the frequency of appearance in different documents in the document collection. This appearance frequency for an entity is represented by a small bar at the left end of each item in the list. In Figure 2(B), the organization list is sorted by frequency, the two other lists are sorted alphabetically.
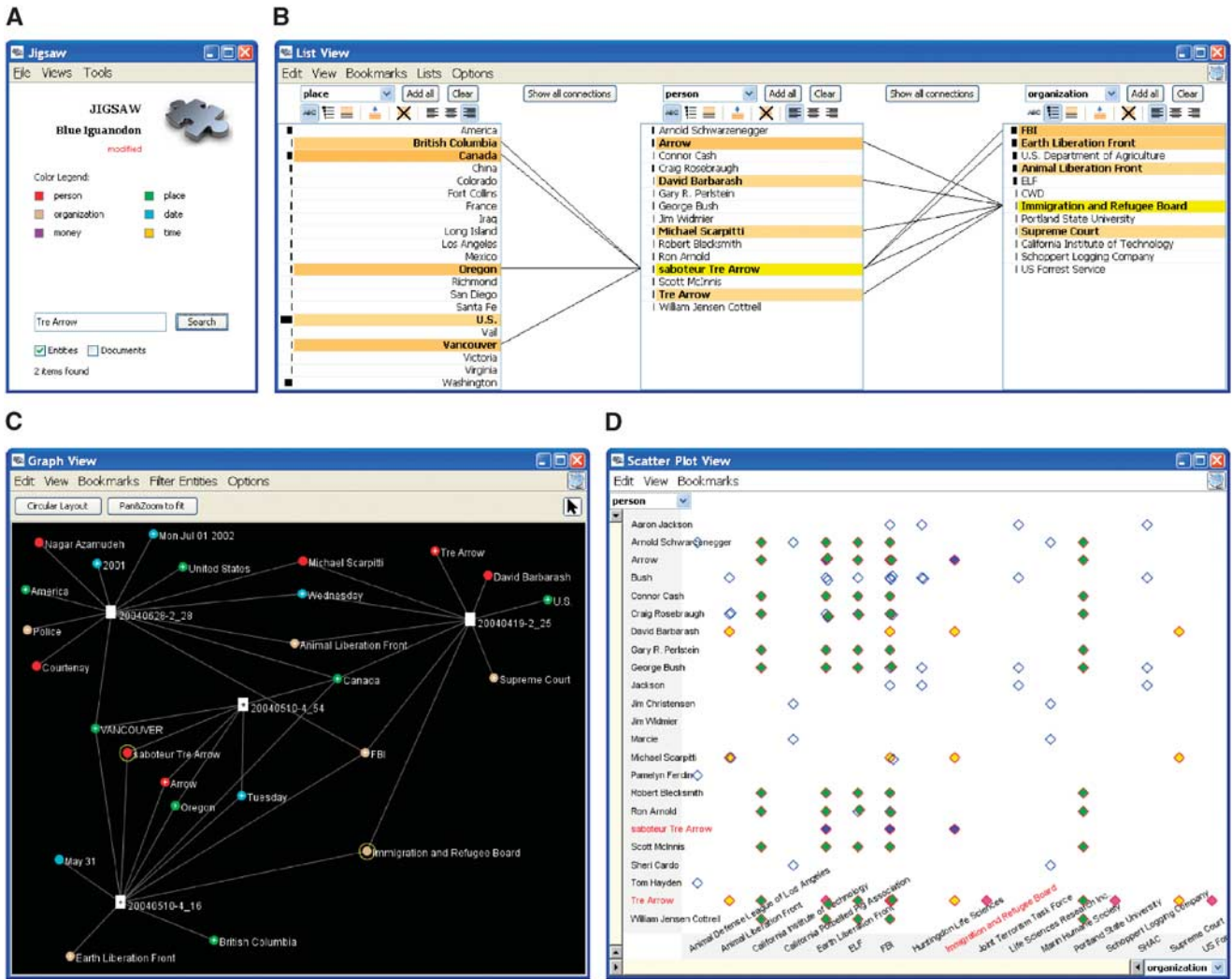
Figure 2 (A) The Control Panel. (B) The List View. Selected entities are shown in yellow and connected entities are indicated by the joining diagonal lines and the orange shading. Darker shading represents stronger connections to the selected entities. (C) The Graph View. Documents are white rectangles and entities are circles colored by type. Edges connect documents to the entities they contain. (D) The Scatter Plot View. Each axis enumerates a list of entities. Diamonds in the center indicate documents containing particular pairs of entities, one from each axis at the relative x, y position.

For some tasks it is useful to consider all entities of a specific type, for example to find out what place has the highest frequency in a document collection. The button 'Add all' allows the user to add all entities of the selected type to a list. If a list of entities is too long for all the items to fit in the view, scrollbars appear to aid navigation. In this situation, many connected items may not be visible in the view. Thus, the List View also provides a mode in which all selected and connected entities are automatically moved to the top of the list.

Graph View

The Graph View, illustrated in Figure 2(C), represents documents and their entities in a traditional node–link

graph/network visualization common in many other systems. Entities are depicted as labeled circles colored according to their type. Documents are represented as white rectangles. The edges from documents to the entities they contain are shown as well.

Following the query-based exploration approach of `Jigsaw`, the Graph View does not automatically draw all the documents and entities as one large network. We thought that a layout of such a large network would be overwhelming and difficult to understand, and thus would not be as helpful to the analyst in our context. Instead, `Jigsaw`'s view is incremental. Display events (triggered from a query or another view) place documents and entities on the display, and then mouse clicks on items can expand or collapse their connections.

The view uses a simple layout algorithm. Both documents and individual entity nodes are randomly positioned in the plane when they are first shown. When a document is being expanded all its entities are displayed as a group, and are drawn at random positions in a small circle around the document like satellites orbiting a planet. When an entity is being expanded, all the documents in which it occurs are displayed in a bigger circle around the entity to create some space for further expansion of those documents. If documents or entities are positioned outside of the visible screen area, the view automatically pans and zooms to make sure that all items are visible.

We have found that this simple layout provides reasonable drawings for Jigsaw's needs. In addition, the user can click on any entity or document and drag it to a new location. To ease this manual re-layout process, dragging a document or an entity while holding down the SHIFT-key also moves all its connected items along.

The entities-as-satellites graph visualization provides another important connections view in Jigsaw since the user can see all the different entities mentioned in a document together. Furthermore, the visualization shows an entity mentioned in multiple documents via the lines drawn from the different documents to that entity. We have found the view to be useful in an interactive exploration mode – the analyst displays an initial document or entity, then expands the item to reveal its relations, and expands one of those items to reveal more, and so on. This type of interaction alternately reveals documents and connected entities.

Multiple nodes can be selected via CTRL-key clicks or by rubber-banding a rectangular or circular selection region. A small plus sign marks nodes that are not fully expanded. The system also provides an inverse selection operation that toggles the selected/unselected state of each node. Other commands allow node(s) to be hidden (they retain the same position if they are subsequently shown again), different types of entities to be filtered from the display, and edges or labels to be removed.

To ease navigation, zooming and panning is implemented through mouse wheel interactions. In addition, the Graph View also provides a 'Circular Layout' command that repositions all the visible documents equidistant around a large circle in the view. Entities connecting to only one document are drawn near that document, but outside the circle. Entities connecting to more than one document are drawn inside the circle. Thus, the set of entities easily noticeable inside the circle represents a more highly connected network of entities that may be related in important ways and likely should be examined more closely. Figure 5 in the section 'Other Application Domains' illustrates the circular layout approach.

Scatter Plot View

The Scatter Plot View, as shown in Figure 2(D), highlights pairwise connections between entities and shows the documents containing the coincidences through a pseudo Starfield display.[27] The user specifies, through a selection box on each axis, the entity type to be placed on that axis. All entity names of that type that have been displayed by queries or display events from other views are then (logically) listed along the axis – in chronological order for dates and in alphabetical order for other entities. If entities from each of the two axes appear together in a document, a diamond is drawn in the view at the conjunction of the two entity's positions along the respective axes. Since a document can contain more than one entity of the same type, multiple visual representations (diamonds) of the same document can appear together in the view at the same time. When moving the mouse over a document, all instances of that document in the view are highlighted. The user can also mark a document with a specific color. This mark will be applied to the document even if the user changes entity types on the axes, thus allowing the user to follow it across other entity-type investigations.

Representative entity labels are drawn in a readable font size at equally spaced intervals along each axis to help the viewer. However, it is likely that many more entities exist in each category than can be shown this way. The view displays these other labels in a tiny illegible font size to provide a hint about the quantity of labels missing. When the user moves the mouse pointer over an entity name, the scatter plot magnifies that item to be readable.

With a large set of entities displayed on the axes, the display area can become cluttered with many diamonds representing relevant documents. To address that problem and help the viewer focus on sets of entities, the Scatter Plot View provides range sliders on each axis so that the viewer can zoom in on a segment of the axis. The view then updates to only show documents containing entities in that smaller range. We have found this capability particularly useful when dates are shown on an axis as a type of time-series view. The viewer can narrow the display to focus on a small interval of time.

Document View

Reading and understanding the actual text documents is a crucial part of an investigative process. To facilitate this need, Jigsaw includes a textual Document View as shown in Figure 3(A). Multiple documents can be stored in one Document View – they are organized in a list at the left border and the user can select a particular document to be displayed. All the entities in the document are highlighted in colors consistent with the color coding of the entity types. The tag cloud at the top of the view describes the contents of the marked documents in the document list.

Documents that already have been viewed are colored blue in the list of available documents and a number at the left side of their list entry shows how often they have been viewed. The list of available documents can be sorted either, alphabetically, by view-count, or it can represent
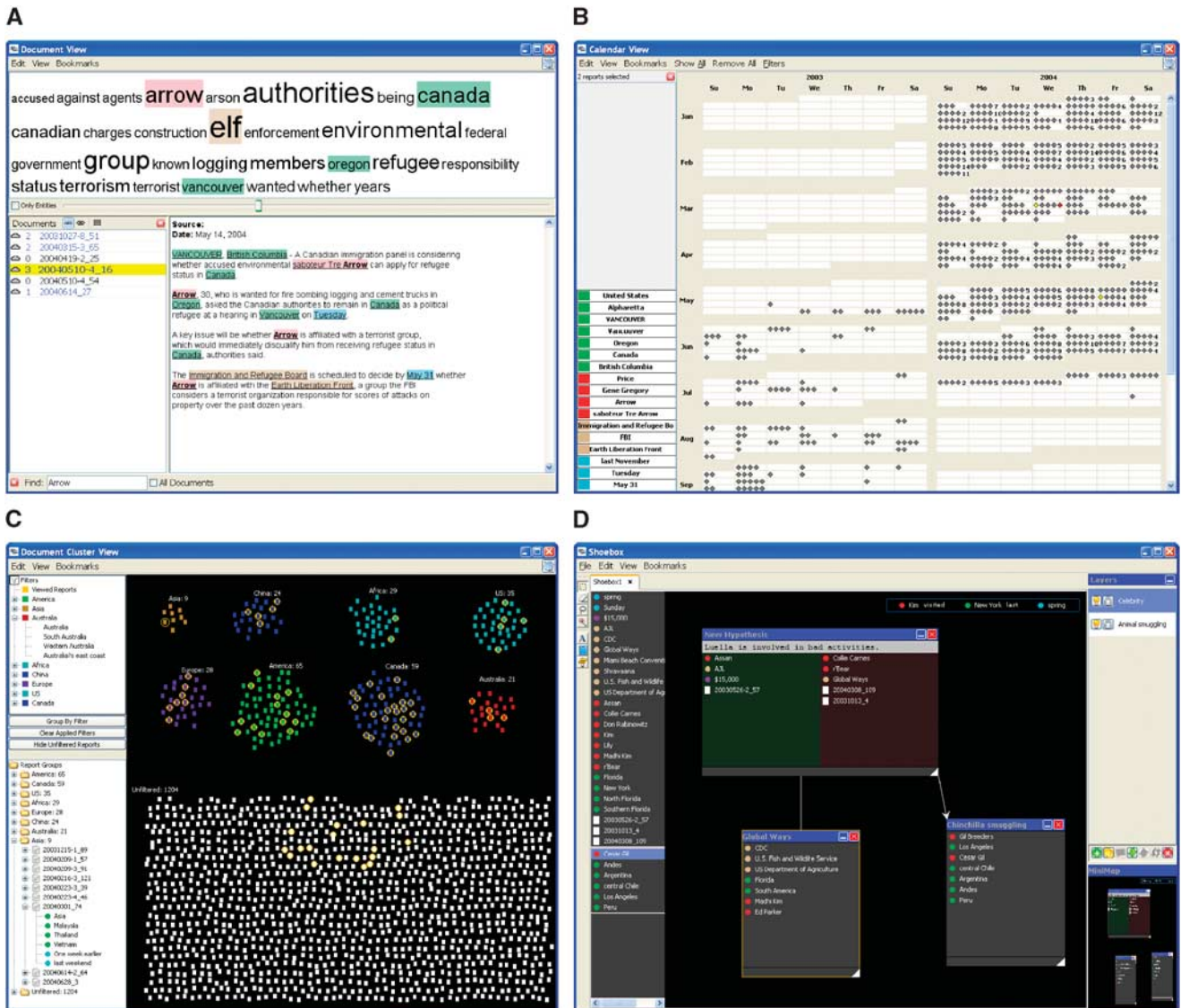
Figure 3   (A) The Document View. The document discussed in the second section is displayed with entities highlighted and colored by type. (B) The Calendar View. Small diamonds represent documents or entities connected to the date in which they appear. (C) The Document Cluster View. Each small rectangle represents a document in the collection, and can be colored and grouped based on entities that do or do not appear in it. (D) The Shoebox. The group, hypothesis, sentences, and link feature are used to organize evidence.

the order in which documents have been added to the view. The find function at the bottom of the view allows the user to quickly find entities or text strings in the open document or in all documents that are currently available in the view.

The automated entity identification process in `Jigsaw` typically is not perfect. Some entities may not be identified at all, some may have an incorrect entity type assigned, and some identified simply may not be entities. The Document View provides functionality to correct such errors. Identified entities can easily be removed or their assigned type changed to another existing type or to a new entity type by right clicking on the identified entity and selecting the relevant command from a pop-up menu. If an entity was missed during the identification process, it can be interactively designated as an entity by selecting the text and assigning an entity type. All these operations to modify entities can be applied to a single document or across all documents in the collection. This allows the user to make local corrections (e.g. the entity 'Paris' is always identified as a location but in one document it is a name) as well as global changes ('BMW' is not identified as an entity and should be added as a new entity type 'car' in all documents).

Calendar View

The Calendar View, shown in Figure 3(B), presents different documents and entities from the data set in the context of a familiar calendar showing years, months, weeks, and days. The small diamond items drawn on a particular day represent documents (colored gray) or entities (colored according to its type) in the context of the date(s) noted in documents in which they appear. When the number of items associated with a day is too large to draw them all in that region, a number is drawn indicating how many others appear on that day but are not shown. As the user moves the mouse over that day, a larger rectangle pops up and shows all the items. When the user moves the mouse pointer over a document-representation diamond drawn in the calendar, all the entities appearing in that document are shown on the left.

The Calendar View is particularly useful to explore patterns over time that are not visible by examining connections within documents. For example, if Peter is dropping off packages on Mondays (mentioned in one document) and Mary is picking up packages on Tuesdays (mentioned in another document) there exists a potential connection between Peter and Mary even though they are never mentioned together in one document. The Calendar View can reveal these types of connections by visualizing the time pattern.

Document Cluster View

The Document Cluster View, shown in Figure 3(C), represents all the documents in the collection, or some subset of the documents, as rectangles. The user can drag individual documents or collections of documents to create different clusters. In addition, all queries and display events triggered by other views are stored in a list of filters (the top left tree in the view) and the user can apply those filters to partition the collection of rectangles into different clusters. In Figure 3(C) the user displayed different places and then clustered the document collection according to those display events. The user can also highlight the documents viewed or not viewed so far, or all the documents that either include or do not include some entity (using the document/entity tree at the bottom left in the view).

The Document Cluster View is helpful to gain an overview of the document collection, to organize the documents according to keywords, to decide what documents could be relevant for a specific task and should be inspected, and what documents are not relevant and should be removed from the document collection.

Shoebox

The Shoebox, shown in Figure 3(D), is not really a document/entity view. Instead, it provides rudimentary evidence marshalling support. The analyst can add relevant entities and documents from other views to the Shoebox; they appear first in the 'inbox-area' on the left side of the Shoebox. These added items can be distributed across different layers, combined to groups or hypotheses and linked together to form sentences. It is also possible to integrate hyperlinks to bookmarks of other views. Provided with these basic functions, the analyst can organize the collected evidence. These sense-making artifacts support the analyst's thinking process in a visual way and begin to provide marshalling support as exhibited by other systems.[28–31]

Scenario

In this section we walk through an analysis scenario with a fictional data set to demonstrate more specifically how Jigsaw supports an analyst. The Jigsaw website[14] provides a video that demonstrates the scenario actions.

Suppose that an analyst received information regarding a suspicious person named Michael Jones. To learn more about him, the analyst starts Jigsaw, opens the data set, displays the List View, selects *Person* as the entity to be shown in the left list, adds all person entities to the list, and sorts the list by frequency. Michael Jones appears at the second position and the long bar next to the name indicates that he is mentioned in a number of documents. In order to explore people associated with Jones, the analyst places *Person* entities in the second list as well and moves the people associated with Jones to the top. The color mappings imply that Martin Clark has the strongest connection to Jones since his name is colored in a dark shade of orange (see Figure 4(A)).

To verify this connection the analyst examines the Document View to read the documents about Michael Jones. He is mentioned together with Martin Clark in two documents (FBI_11 and FBI_35) and thus the connection seems plausible.

The analyst returns to the List View, selects both Clark and Jones, and puts *Organization* entities into a third list which reveals that both men have connections to the same organizations. The Revolution Now Scholarship Fund has the strongest connections of any organization, so the analyst continues the exploration on it.

The Document View shows two documents mentioning the Scholarship Fund. Report FBI_35 mentions that Michael Jones donated $48,000 to the fund on the stipulation that the donation be equally split among six students, Martin Clark being one of them. The analyst also notes that the six students form three pairs – where students in each pair live close to each other. This raises suspicions that the students might be collaborating.

Proceeding, the analyst brings up the three documents about Martin Clark and William Brown (who both live in Virginia) by displaying them in the List View. Two of the documents were already encountered in this investigation and the third, FBI_41 (see Figure 4(B)), states that a month ago Clark and Brown took a cruise together from Hampton to Kingston, Jamaica. Furthermore, both are again on this cruise right now. The document also says
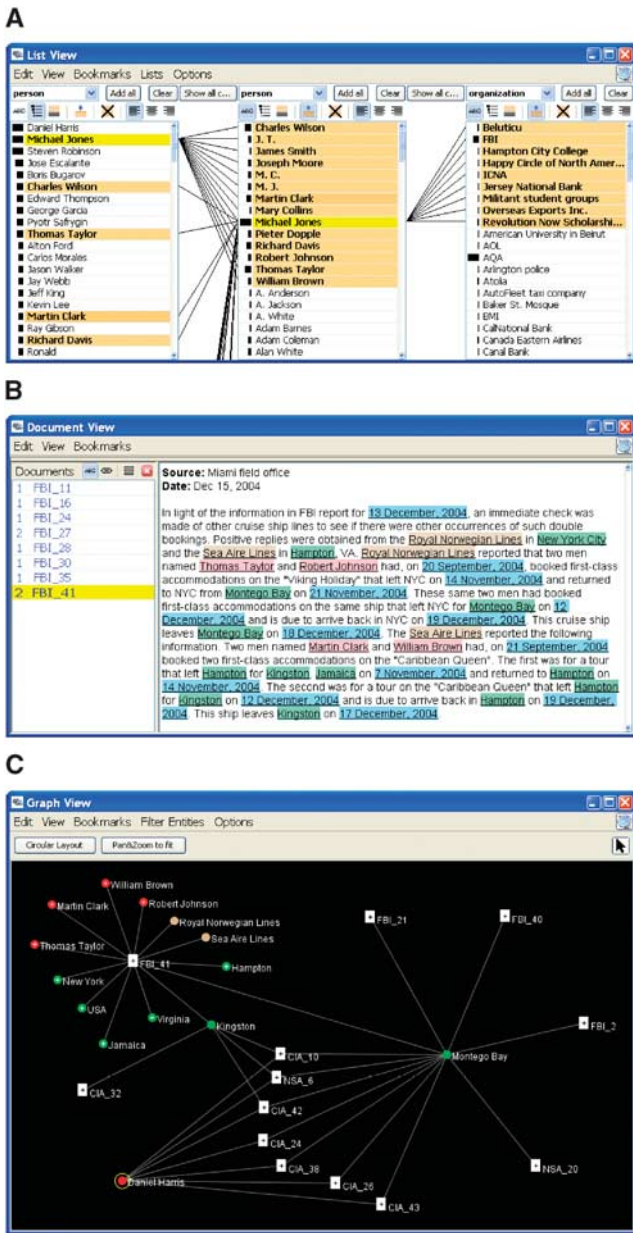
Figure 4   (A) The List View, (B) Document View, and (C) Graph View from the example scenario.

analyst selects and reads these documents in the Document View.

All three documents mention the person Daniel Harris who works in Montego Bay. The analyst issues a query on Harris, showing the man's entity in the Graph View. She expands his node and connections to seven more documents show up (see Figure 4(C)). Upon reading these documents, the analyst learns that Harris traveled from Montego Bay to Kingston on 1 December and passed a package to a person named Edward Thompson.

The analyst concludes the investigation hypothesizing that suspicious activities are planned involving some of these individuals traveling on cruise ships in the Caribbean and with potential packages of interest. The analyst suggests that further investigation be conducted focusing on Daniel Harris and related activities.

VAST 2007 contest

We participated in the VAST 2007 Contest[16] using `Jigsaw` as an analytic aid, and we won the university division of the competition.[32,33] While working on our contest entry, we learned a great deal about how to effectively use the system, we fixed usability problems, and we implemented more functionality that was useful for specific tasks. The system matured quite a bit during that process. One key finding was that it is crucial for `Jigsaw` to have correctly identified entities since they define the connections across the documents. If entities are missed or wrongly identified, connections are missing or misleading. We added the entity modification functionality to the Document View to address that important problem.

We also were invited to participate in a special live contest at the conference. We worked with a professional analyst on a smaller data set using `Jigsaw`. Feedback from the analyst was very beneficial. He enjoyed working with `Jigsaw`, in particular the circular layout function of the Graph View, since it facilitates the visual exploration process and shows at a glance what entities a set of documents has in common. Some other points for improving `Jigsaw`'s capabilities arose as well: adding a view to visualize geographic information and providing better support to build a timeline in the Shoebox would be helpful for an analyst.

Other application domains

So far, we have concentrated on applying `Jigsaw` in domains like intelligence and law enforcement but we can envision using it in a more general sense-making process and in other domain areas as well.

We already applied `Jigsaw` to explore connections between people, places, and organizations in the Bible. We built a set of documents by dividing the Bible into its chapters, imported that data set into `Jigsaw`, and investigated the connections. The Graph View in Figure 5 shows the social network of 'Cain' and 'Abel' as a circular

that two other scholarship recipients, Thomas Taylor and Robert Johnson, took a cruise together from New York City to Montego Bay last month and they are also currently on this cruise again.

Now, to gain a deeper understanding of the connections between the people and places, the analyst examines report FBI_41 in the Graph View. After expanding the node for the report and filtering out the date entities, the analyst expands the nodes representing Kingston and Montego Bay. The view reveals that both are connected to three document nodes: FBI_14, CIA_10, and NSA_6. The

Figure 5 The Graph View showing Cain's and Abel's social network in the Bible along with the places they are associated with.

layout along with places mentioned in chapters in which they appear.

One area that seems well-matched to Jigsaw's capabilities is academic research involving journal and conference papers. People new to a discipline may want to learn about the important researchers and papers in the area. Existing tools like PaperLens[34] address such scenarios already but trails of connected articles are not so easy to follow. In Jigsaw we can use metadata of papers such as authors, institutions, categories, etc., as entity types and then visualize and explore connections between these entities. A second scenario is to define entity types relevant for a specific research area. For example, in biology papers about research on genes, the different genes could map to a set of entities and biologists could benefit by learning about connections of genes across all relevant publications.

Another possibility would be to use Jigsaw to explore online articles such as web news reports or blogs on topics like political debates. Using the import mechanism of Jigsaw, a lay user could build a set of loosely related documents about a particular topic or person (retrieved using a search engine), and then use Jigsaw to find connections between entities among the documents. That would help to filter the collection to a smaller number of documents for careful reading.

In another paper[35] we discuss in more detail how to use Jigsaw in a human-centered sense-making process using visual exploration.

Related work

A growing number of research and commercial systems are using visualization and visual analytic techniques to help support investigative analysis.

Analyst's Notebook from i2 Inc.[36] provides a semantic graph visualization to assist analysts with investigations. Nodes in the graph are entities of semantic data types such as person, event, organization, bank account, etc. While the system can import text files and do automatic layout, its primary application appears to be helping analysts in manually creating and refining case charts.

Oculus Info Inc. provides a suite of systems for different aspects of investigative analysis. GeoTime[37] visualizes the spatial inter-connectedness of information over time overlaid onto a geographical substrate. It uses an interactive 3D view to visualize and track events, objects, and activities both temporally and geospatially. TRIST[38] allows analysts to formulate, refine, organize and execute queries over large document collections. Its user interface is a multi-pane view that provides different perspectives on search results including clustering, trend analysis, comparisons, and difference. Information retrieved through TRIST then can be loaded into the SANDBOX system,[31] an analytical sense-making environment that helps to sort, organize, and analyze large amounts of data. The system's goal is to amplify human's insights with computational linguistic, analytical functions, and by encouraging the analyst to make thinking more explicit. The system offers interactive visualization techniques including gestures for placing, moving, and grouping information, as well as templates for building visual models of information and visual assessment of evidence. An evaluation experiment of the SANDBOX system showed that analysts using the system did higher quality analysis in less time than using standard tools. Jigsaw provides a different style of visual representation of document entity data to analysts; TRIST and SANDBOX provide more authoring and organizational infrastructure.

The COPLINK system[15,39] and related suite of tools was developed to help law enforcement officials more easily extract information from police case reports and analyze criminal networks. The system uses data mining techniques to set up a concept space of entities and objects that can be searched to find related items. Visualization support consists of a hyperbolic tree view and a spring-embedder graph layout of relevant entities. Jigsaw's Graph View shares the exploratory sense of the hyperbolic tree but does it for a connected graph. Beyond that, Jigsaw provides a more varied suite of visualizations of documents and their entities, but it does not presently include the sophisticated mining and analytic capabilities of COPLINK.

IN-SPIRE[17] is a system for exploring textual data in document collections and it has been used extensively for intelligence analysis.[40] The system generates a 'topical landscape', either through a 3D surface plot or a galaxy-style view, that supports queries, provides the

possibility to analyze trends over time, and allows analysts to discover hidden information relationships among documents. Its goal is to identify and communicate the different topics and themes, and then allow the analyst to inspect the documents more deeply through interactive analysis. Jigsaw differs in its focus on exploring relationships among the entities in documents.

ENTITY WORKSPACE[28] is a tool to amplify the usefulness of a traditional evidence file that is widely used by analysts to keep track of facts. It provides an explicit model of important entities to help the analyst to find and re-find facts rapidly, discover connections and identify important documents and entities to continue the exploration. The system is just one of a suite of tools from PARC directed at assisting sense-making.[41]

The two systems probably closest to our work, KANI and ProSPECT, come from PNNL. KANI,[29] like Jigsaw focuses on visualizing entities from textual documents. KANI has two main views, a document viewer that highlights entities and their selected relationships and a graph view that shows different entities connected in a node–link structure. The system provides extensive filtering capabilities to the analyst and includes automated associate components that help with activities like hypothesis refinement and assumption testing. Jigsaw goes beyond KANI in the variety and style of the interactive visualizations provided, but KANI has a more complete infrastructure for supporting reasoning and hypothesis formulation.

Sanfillipo and colleagues at PNNL[30,42] created the ProSPECT system that extracts scenario information from unstructured intelligence data sources. Their system was built to provide multiple views on multiple monitors as does Jigsaw. More specifically, they provide three main views or spaces: a source space for collecting documents and sources; a marshalling space for examining evidence and reasoning; an analysis space for constructing hypotheses. ProSPECT performs language analysis and uses ontologies to facilitate reasoning, and it supports analysts to put confidence levels on evidence and hypotheses. Jigsaw provides a richer set of document and entity views, with a focus on entity connection, than does ProSPECT, but the source intake and hypothesis exploration facilities of ProSPECT go beyond what Jigsaw currently provides.

Our approach differs primarily from that of the systems described above in its focus on visually representing connections and relationships between entities in document collections. We advocate a multitude of data views with simple, consistent user interactions for promoting fast exploration and discovery. Furthermore, Jigsaw provides a system model where user interaction is a first-class object, helping to expose the entity connections, and providing for easier extensions to new styles of views.

Discussion and future directions

While Jigsaw provides a number of capabilities that we believe will be useful for investigative analysis, our work has only begun to scratch the surface of what is possible in this area. Numerous avenues of research and extensions to the system are possible in future work. In fact, we have many already underway.

Because the system has yet to be rigorously evaluated, that is an obvious next step. We have conducted initial usability assessments of the individual system views to improve each's effectiveness and we are presently conducting studies to compare groups of people (in this case, graduate students) performing document analysis both with and without Jigsaw present. Obviously, trial use of the system and evaluation by professional analysts is a key future action to understand Jigsaw's strengths and weaknesses and to reflect on its utility. Our use of the system in the VAST 2007 Contest provides at least a little empirical evidence that it is useful in the kinds of investigative scenarios exemplified by the contest.

Jigsaw currently does not 'start cold' well. By this, we mean that Jigsaw does not provide an overview of the key themes and trends across a document collection to help an analyst begin exploration as is done well by, for instance, the IN-SPIRE[17] system. Jigsaw is more useful when an analyst has a few target entities of interest and begins an investigation exploring them. The document analysis and clustering capabilities of IN-SPIRE recently have been exported into web services, however, and we are planning to incorporate these capabilities into Jigsaw. Additionally, we are exploring more 'discount' methods of providing topical overviews and key themes in the documents such as the present use of tag clouds to present important words used in and across the documents.

Presently, Jigsaw provides a higher level of visualization support than automated analytic support. In an idealized visual analytics system, these two aspects are each deeply present and they are blended seamlessly. As discussed earlier in the article, enhancing the computational linguistic support of the system could provide powerful new capabilities. Furthermore, analytic techniques from the link analysis community could provide more sophisticated graph analytic capabilities to Jigsaw.

Jigsaw has been used to examine document collections numbering in the thousands and we believe that it could handle collections in the tens of thousands reasonably well, but much larger document collections clearly may be arise, particularly if web searches are performed. Scalability is always an issue in any kind of visual analytic system. In order to address larger collections, we have considered notions such as a document 'holding area' where documents are available for analysis but not fully integrated into the system views yet. Analysts should be able to fluidly add and dismiss documents without affecting system performance or utility.

Jigsaw's model of entity connection being calculated at the document level is both simple and useful, but as document sizes increase, this model becomes less accurate and informative. We plan to adopt a more fine-grained model by classifying connections according to their locality: entities are loosely connected if they appear in the same

document, they are moderately connected if they appear in the same paragraph, and they are strongly connected if they appear in the same sentence. Determining entity connections on a semantic level much like that done by Sanfilippo[30] may be needed in order to facilitate analysis on larger documents.

Our use of the system has suggested the need for a dedicated geospatial view. The `GeoTime` system[37] provides a good example here. In sample analysis sessions, we have noted the absence of explicit views supporting geolocation-related analysis. Of course, adding even more types of views raises issues concerning the multiplicity of views – could an abundance of representations overwhelm analysts rather than assist them? How many different kinds of views can profitably be used together?

Unstructured text documents are currently the analytic focus of `Jigsaw`, but many other types of data and evidence exist. For instance, structured data (often in the form of spreadsheets or databases) might include air travel records, vehicle registrations, bank account transactions, immigration records, cargo container logs, etc. By expanding our system to also read structured data, we believe that its analytic capabilities will be greatly amplified. The key research question here is how to integrate structured data with the current document and entity (type) model of `Jigsaw`. How are rows and columns (objects and attributes) interpreted? Analysts should not be forced to specify the mapping of each field of a database to add its data to an investigation.

In the views presented by `Jigsaw`, entity connections to documents and other entities are taken as facts or truth, as is commonly done in visualization.[43] System views illustrate these connections using the straightforward model of connection employed now. We have learned through informal discussions with investigative analysts that reports often may come from sources with varying levels of reliability or the reliability of a document simply may not be known. Furthermore, reports may describe events that are uncertain.

For instance, a police report may come from a source that is unreliable, thus raising questions about the activities described in the report. Alternatively, intelligence reports may include language such as 'It is probable that…', 'We believe that…', or 'There is little chance that…'. Clearly, to analyze reports like these, relatively sophisticated linguistic analysis is needed and even then, determining the probability of the event is challenging and debatable.[8] Such analysis is not our expertise nor is it a focus of our work, but other researchers do address issues like these. A future challenge for us is to develop representations of the uncertainty and reliability of document information.

Amplifying `Jigsaw`'s support for collaborative analysis is another potential future direction. Our experience working with the system has shown that two analysts can work effectively together at one local machine if sufficient screen space is available, for example, as exists in our lab configuration shown in Figure 1. If more

than two analysts collaborate, even this configuration becomes too restrictive as the bottleneck of only one keyboard and mouse surfaces. Other information visualization researchers have proposed different approaches for scenarios like these. One involves a multi-touch table top computer where multiple people can interact with the display at the same time.[44] This approach has the disadvantage that the types of interactions presently supported are restricted. An alternative approach would be to use a large, hi-resolution wall-style display[45] with personal interaction devices for each analyst. Furthermore, `Jigsaw` could support *distributed* synchronous collaboration so that multiple analysts at different locations could work together on an investigation. `Jigsaw`'s system architecture and event passing actually makes this type of collaborative scenario relatively feasible without extensive work.

Trial exposure to and use of the system by professional analysts uncovered a desire for better review capabilities of the investigation state and history. Because investigations often involve a large amount of data and can be carried out over weeks, months, or even years, it is crucial for analysts to know what part of the data they already have or have not considered and when data first became available. For instance, analysts may have elaborated a hypothesis but then new evidence is found that refutes this hypothesis. Similarly, a particular angle of investigation may not prove fruitful, so the analyst needs to back up and resume exploration from some earlier state. To be able to quickly develop an alternate model or view, it is crucial for analysts to know what data they had already considered, what data led to that hypothesis, what other hypotheses were already considered and discarded, what leads are still open, and what data has not yet been considered at all.

The need for better tools to augment the process of documenting and presenting the results of analysis, so-called production, presentation, and dissemination, has been identified.[5] Views in `Jigsaw` can be bookmarked and linked in evidence marshalling, but better support for annotation to provide visual summaries of the evidence used to reach actionable conclusions is needed.

Finally, one could use capabilities like those in `Jigsaw` as a general-purpose means to explore search engine results. An environment incorporating a powerful search engine and `Jigsaw`'s exploration facilities could be a helpful sense-making aid.

Conclusion

Every day investigative analysts are faced with the challenging task of assessing and making sense of large bodies of information. Technological aids that promote data exploration and augment investigators' analytical reasoning capabilities hold promise as one way of assisting analysis activities.[5,31,41] In a workshop of intelligence analysis professionals,[46] working groups generated a list of the top 10 needs for intelligence analysis tool

development. One item was 'Dynamic Data Processing and Visualization' that was further elaborated as follows:

'Solutions are needed that transcend what is typically described as 'visualization' – in contrast to a predominantly 'passive' relationship between the system that displays complex visualizations and the analyst who still must digest and interpret them. What is needed is a much more interactive and dynamic relationship in which the analyst is better able to explore the information within the visualization.'

Herein we present `Jigsaw`, a system designed to assist analysts with foraging and sense-making activities across collections of textual reports in just this manner. `Jigsaw` presents a suite of views that highlight connections between entities within the reports. Through interactive exploration, analysts are able to browse the entities and connections to help form mental models about the plans and activities suggested by the report data.

`Jigsaw` is not a substitute for careful analysis of the reports, however. Instead, it acts as a visual index that presents entity relations and links in forms that are more easily perceived, thus suggesting relevant reports to examine next. Other systems sometimes put too much information into a single complex view, with the result that though information may be present, it is harder to discern and is much less flexible from the analyst's viewpoint. Our approach hinges on multiple, easy-to-understand views with simple, clear interactions. In creating the visualizations we leveraged well-known representations from the field of information visualization and augmented them with interactive operations useful for showing connections between entities.

`Jigsaw` provides four main contributions to the growing area of visual analytics for investigative analysis:

- A suite of interconnected, cooperating, interactive visualizations that present multiple perspectives on a document collection and, in particular, connections between entities across the collection.
- Advocacy and illustration of the value of large pixel space approaches to investigative analysis where multiple computer displays enable richer portrayals of the documents and data being examined.
- A fundamentally simple user interaction model in views (single-click is select and double-click is expand/collapse) in order to make the system less complex for analysts to use.
- Synthesis of all these capabilities into a working system that has already exhibited investigative value and that holds potential for further use by professional analysts.

Acknowledgements

References

1 Russell DM, Stefik MJ, Pirolli P, Card SK. The cost structure of sensemaking. *CHI '93: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, Netherlands) ACM: New York, NY, USA, 1993; 269–276.

2 Klein G, Moon B, Hoffman R. Making sense of sensemaking 1: alternative perspectives. *IEEE Intelligent Systems* 2006; **21**: 70–73.

3 Norman D. Visual representations. In *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Addison-Wesley: Reading, MA, 1994.

4 Card SK, Mackinlay J, Shneiderman B. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann: Los Altos, CA, 1999.

5 Thomas JJ, Cook KA. *Illuminating the Path*. IEEE Computer Society: Chicago, 2005.

6 Wong PC, Thomas J. Visual analytics. *IEEE Computer Graphics and Applications* 2004; **24**: 20–21.

7 Pirolli P, Card S. Sensemaking processes of intelligence analysts and possible leverage points as identified through cognitive task analysis. *Proceedings of the 2005 International Conference on Intelligence Analysis* (McLean, VA). https://analysis.mitre.org/proceedings/index.html, May 2005.

8 Heuer R. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, Central Intelligence Agency: Washington, DC, 1999.

9 Johnston R. *Analytic Culture in the United States Intelligence Community*: *An Ethnographic Study*. Central Intelligence Agency Washington, DC, 2005.

10 Amar R, Eagan J, Stasko J. Low-level components of analytic activity in information visualization. In *Proceedings of IEEE Symposium on Information Visualization (InfoVis '05)* (Minneapolis, MN), IEEE Computer Society Press: Siver Spring, MD, October 2005; 111–117.

11 Amar R, Stasko J. Knowledge precepts for design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics* 2005; **11**: 432–442.

12 Yi JS, Kang YA, Stasko J, Jacko J. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics* 2007; **13**: 1224–1231.

13 Stasko J, Görg C, Liu Z, Singhal K. Supporting investigative analysis through interactive visualization. *Proceedings of IEEE Symposium on Visual Analytics, Science, and Technology (VAST'07)* (Sacramento, CA), IEEE Computer Society Press: Silver Spring, MD, 2007; 131–138.

14 Jigsaw website. http://www.cc.gatech.edu/gvu/ii/jigsaw (accessed 21 March 2008).

15 Chen H, Zeng D, Atabaksh H, Wyzga W, Schroder J. Coplink: managing law enforcement data and knowledge. *Communications of the ACM* 2003; **46**: 28–34.

16 IEEE VAST 2007 CONTEST. http://www.cs.umd.edu/hcil/VASTcontest07 (accessed 21 March 2008).

17 Wise JA, Thomas JJ, Pennock K, Lantrip D, Pottier M, Schur A, Crow V. Visualizing the non-visual: spatial analysis and interaction with information from text documents. *Proceedings of IEEE Symposium on Information Visualization (InfoVis'95)* (Atlanta, GA), IEEE Computer Society Press: Silver Spring, MD, October 1995; 51–58.

18 Chau M, Xu J, Chen H. Extracting meaningful entities from police narrative reports. *2nd National Conference on Digital Government Research* (Los Angeles, CA), 2002.

19 Jackson P, Moulinier I. *Natural Language Processing for Online Applications: Text Retrieval, Extraction, and Categorization*. John Benjamins: Amsterdam, Netherlands, 2002.

20 GATE – General Architecture for Text Engineering. http://gate.ac.uk/ (accessed 21 March 2008).

21 Balie. http://balie.sourceforge.net/ (accessed 21 March 2008).

22 FreeLing. http://garraf.epsevg.upc.es/freeling/ (accessed 21 March 2008).

23 Connexor. http://www.connexor.eu/ (accessed 21 March 2008).

24 Cicero. Language Computer Corp. http://www.languagecomputer.com/ (accessed 21 March 2008).

25 Sanfilippo A, Franklin L, Tratz S, Danielson G, Mileson N, Riensche R, McGrath L. Automating frame analysis. *First International Workshop on Social Computing, Behavioral Modeling, and Prediction* (Phoenix, AZ), Springer Science+Business Media: New York, NY, April 2008; 239–248.

26 Hughes FJ. Discovery, proof, choice: the art and science of the process of intelligence analysis, case study 6, 'All Fall Down'. Unpublished report, 2005.

27 Ahlberg C, Shneiderman B. Visual information seeking: tight coupling of dynamic query filters with starfield displays. *Proceedings of ACM CHI '94*, (Boston, MA), ACM: New York, NY, April 1994; 313–317.

28 Bier EA, Ishak EW, Chi E. Entity workspace: an evidence file that aids memory, inference, and reading. *Proceedings of the IEEE International Conference on Intelligence and Security Analysis (ISI '06)*, (San Diego, CA), Springer-Verlag: Berlin, May 2006; 466–472.

29 Chappell AR, Cowell AJ, Thurman DA, Thomson JR. Supporting mutual understanding in a visual dialogue between analyst and computer. *Proceedings of Human Factors and Ergonomic Society 48th Annual Meeting* (New Orleans, LA), Human Factors and Ergonomics Society: Santa Monica, CA, September 2004; 376–380.

30 Sanfilippo A, Baddeley B, Cowell AJ, Gregory ML, Hohimer R, Tratz S. Building a human information discourse interface to uncover scenario content. *Proceedings of the 2005 International Conference on Intelligence Analysis* (McLean, VA), https://analysis.mitre.org/proceedings/index.html, May 2005.

31 Wright W, Schroh D, Proulx P, Skaburskis A, Cort B. The sandbox for analysis: Concepts and methods. *Proceedings of ACM CHI '06* (Montreal, Quebec), ACM: New York, NY, April 2006; 801–810.

32 Görg C, Liu Z, Parekh N, Singhal K, Stasko J. Jigsaw meets Blue Iguanodon – The VAST 2007 Contest. *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST '07)* (Sacramento, CA), IEEE Computer Press: Silver Spring, MD, October 2007; 201–202.

33 Plaisant C, Grinstein G, Scholtz J, Whiting M, O'Connell T, Laskowski S, Chien L, Wright TAW, Görg C, Liu Z, Parekh N, Singhal K, Stasko J. Evaluating visual analytics: the 2007 visual analytics science and technology symposium contest. *IEEE Computer Graphics and Applications* 2008; **28**: 12–21.

34 Lee B, Czerwinski M, Robertson G, Bederson BB. Understanding research trends in conferences using PaperLens. *CHI '05: CHI '05 Extended Abstracts on Human Factors in Computing Systems, Proceedings of ACM CHI '05* (Portland, OR), ACM: New York, NY, 2005; 1969–1972.

35 Stasko J, Görg C, Liu Z. Sensemaking across text documents: human-centered, visual exploration with Jigsaw. *Sensemaking Workshop @ CHI 2008* (Florence, Italy), http://dmrussell.googlepages.com/sensemakingworkshopchi2008, April 2008.

36 i2 – Analyst's Notebook. http://www.i2inc.com/ (accessed 21 March 2008).

37 Kapler T, Wright W. Geotime information visualization. *Information Visualization* 2005; **4**: 136–146.

38 Jonker D, Wright W, Schroh D, Proulx P, Cort B. Information triage with TRIST. *Proceedings of the 2005 International Conference on Intelligence Analysis* (McLean, VA) https://analysis.mitre.org/proceedings/index.html, May 2005.

39 Xu J, Chen H. Criminal network analysis and visualization. *Communications of the ACM* 2005; **48**: 101–107.

40 Hetzler E, Turner A. Analysis experiences using information visualization. *IEEE Computer Graphics and Applications* 2004; **24**: 22–26.

41 Card SK. Leverage points and tools for aiding intelligence analysis. Unpublished report presented at the 2007 Human–Computer Interaction Consortium, 2007.

42 Sanfilippo A, Baddeley B, Posse C, Whitney P. A layered dempster-shafer approach to scenario construction and analysis. *Proceedings of the IEEE International Conference on Intelligence and Security Analysis (ISI '07)* (New Brunswick, NJ), Springer-Verlag: Berlin, May 2007; 95–102.

43 Finger R, Bisantz A. Utilizing graphical formats to convey uncertainty in a decision-making task. *Theoretical Issues in Ergonomic Science* 2002; **39**: 1–25.

44 Isenberg P, Carpendale S. Interactive tree comparison for co-located collaborative information visualization. *IEEE Transactions on Visualization and Computer Graphics* 2007; **13**: 1232–1239.

45 Ball R, North C, Bowman D. Move to improve: promoting physical navigation to increase user performance with large displays. *Proceedings of ACM CHI '07* (San Jose, CA), ACM: New York, NY, April 2007; 191–200.

46 Badalamente RV, Greitzer FL. Top ten needs for intelligence analysis tool development. *Proceedings of the 2005 International Conference on Intelligence Analysis* (McLean, VA), https://analysis.mitre.org/proceedings/index.html, May 2008.