INVISQUE as a Tool for Intelligence Analysis: The Construction of Explanatory Narratives

Chris Rooney, Simon Attfield, B. L. William Wong & Sharmin Choudhury

March 19, 2014

1 Abstract

We report an exploratory user-study in which a group of civil servants with experience of, or involvement in, intelligence analysis used the tool INVISQUE to address a problem using the 2011 VAST dataset. INVISQUE uses a visual metaphor that combines searching, clustering and sorting of document surrogates with free-form manipulation on an infinite canvas. We were interested in exposing the behaviours and related cognitive strategies that users would employ to better understand how this and similar environments might better support intelligence type work. Our results include the observation that the search and spatial features of the system supported participants in establishing, elaborating and systematically evaluating explanatory narratives that accounted for the data. Also, visual persistence at the interface allowed them to keep track of searches and to re-find documents when their importance became apparent. We conclude with reflections on our findings and propose a set of guidelines for developing systems that support sensemaking.

2 Introduction

In seeking to make assessments about the data they investigate, intelligence analysts face great challenges in terms of the complexity of the issues that they address and demands for timeliness and volume of production (Heuer, 1999). Most accounts of sensemaking describe it as an ad-hoc and opportunistic process which involves both bottom-up information discovery, and the top-down application of theories, hypotheses and questions. Despite this, systems tend to focus on either information discovery or the representation of reasoning products, but not both, and those that do tend to separate associated functionality into different windows (Wright, Schroh, Proulx, Skaburskis, & Cort, 2006; Billman & Bier, 2007). We assume that this may impose unnecessary constraints on sensemaking, since information discovery and reasoning are so intimately related.

We report an exploratory study in which we observed an intelligence analysis task being performed by domain experts using a system called INVISQUE. Our interest was to gain insight into design considerations for this type of task. We did this by performing an exploratory cognitive task analysis with the aims of (i) exposing the cognitive strategies that users might employ when using the system to perform an intelligence analysis task, (ii) understanding how system features might be appropriated in the service of, or present an obstacle to, the task.

The system adopts a metaphor in which multiple searches are conducted on an infinite visual canvas and where search results appear as ‘index cards’ that can be either manually sorted or automatically arranged according to meta-data parameters. In this sense, INVISQUE allows the user to seamlessly combine information seeking with manual and automated information structuring in a single interface. As part of the work in this paper, INVISQUE been expanded to embed its search and retrieval processes within an environment that imposes minimal constraints on user-determined visual structuring.

We set a problem using the dataset from mini-challenge three of the 2011 VAST Challenge (Maryland, 2011). We recruited two groups of participants: a small team involved in capabilities development for intelligence analysts and a group of defence researchers with prior experience as intelligence analysis desk officers. Having no prior experience of the dataset we asked them to address an investigation problem using the system. We report on observational data gathered during the task in the context of the objectives mentioned above.

In the next section we review some related work. In section three we discuss two models of sensemaking that influenced how we interpreted our study results and in section four we describe the INVISQUE system. In section five we describe the user-study and its findings, and in section six we discuss the implications of our results.

3 Related Work

Like many sensemaking tasks conducted in electronic environments, intelligence analysis involves a complex interaction between filtering, extracting and structuring large amounts of information, and developing and communicating interpretations to account for the data as well as predictions for the future (Pirolli & Card, 2005). It has been argued that, for information intensive sensemaking, more information is not necessarily a helpful thing (Klein, Phillips, Rall, & Peluso, 2006), rather that greater support is required for evaluating and interpreting data (Heuer, 1999).

During sensemaking, there is a reflexivity that occurs between data, on the one hand, and the evolving products of reasoning (e.g. theories, hypotheses and questions) on the other (Weick, 1995; Pirolli & Card, 2005; Klein et al., 2006; Attfield & Blandford, 2011). For example, data might provoke the construction of an explanatory theory which itself stimulates new questions and new requirements for data. Further, reasoning products may act as organising principles for both information seeking and information structuring. Information visualisation can be used to support both sides of this equation, and yet systems are seldom designed to integrate both kinds of capability. We propose that this may represent a barrier to closely coupled cycles of interaction. In the following we briefly characterise and distinguish the two approaches by example, using the terms system-structured spaces and user-structured spaces.

System-structured spaces allow a user to specify views on some data in terms of properties of that data. The system may use complex algorithms to extract properties and construct a display.

In this category we place systems traditionally classed as ‘visualisations’. In the field of document analysis, examples of system-structured spaces include scatter plots (Ahlberg & Shneiderman, 1994; Kang, Grg, & Stasko, 2010), clustering based on latent semantic analysis or similar (Russell, Slaney, Qu, & Houston, 2006; Gregory, Payne, McColgin, Cramer, & Love, 2007; Kang et al., 2010), and 3D data terrains (Wise et al., 1995). Interaction typically involves filtering large datasets (Yeh & Liu, 2011) to reveal trends and relationships that are within the data or are found through the application of analytics.

User-structured spaces enable the user to freely manipulate entities such that they define relationships rather than discover them. These are tools for thinking and reflecting back the results of that thinking. Significantly, such systems can be used to capture interpretations and ideas that a user places on some data. These ideas extend beyond what is in the data per se, or what can be computationally inferred from it. Interaction might involve moving, grouping or otherwise relating data objects, resulting in structures that may correspond to, for example, themes, arguments or narratives. Such systems leverage a relationship between the way we reason and our use of space. For example, Andrews et al. (Andrews, Endert, & North, 2010) investigated the use of large Powerwall displays, which offer vast screen real estate, as tools for performing sensemaking tasks. They found that users encoded meaning in the spatial relationships between data, such as grouping relevant documents, and keeping important documents closer to the central focal region. This research led to the development of the Analyst’s Workspace (Andrews & North, 2012), which enables analysts to define relationships by freely arranging documents on the workspace, as well as imposing a visual representation of automatically extracted relationships between entities (e.g., identifying the set of documents that mention a particular individual).

Spatial hypertext systems are an early example of how space can support the analytical process (Marshall & Shipman, 1997; Shipman, Hsieh, Maloor, & Moore, 2001). These systems allow analysts to encode relationships between documents in terms of spatial groupings, for example, for supporting document triage. This concept has evolved to support the cognitive processes involved in sensemaking. Sandbox, for example, is a thinking environment for reasoning about data and creating hypotheses (Wright et al., 2006). Analysts are able to drag data onto a canvas, where they can create a spatial map of their thought processes. They can also annotate the canvas with thoughts and theories. Similarly, the Entity Workspace provides an evidence panel for analysis to place relevant documents and generate hypotheses (Billman & Bier, 2007), whilst Dynapad offers a workspace for spatially organising personal document collections (Bauer, Fastrez, & Hollan, 2005). Finally, the S3 system embeds cognition into a visual interface for developing and analysing hypotheses (Ntuen, Park, & Gwang-Myung, 2010).

Few systems combine system-structuring and user-structuring capabilities. However, an exception can be found in ForceSPIRE (Endert, Fiaux, & North, 2012b), which exploits user structuring to inform underlying analytics in a closely coupled mixed initiative interaction - what the authors refer to as *semantic interaction*.

Whilst interaction with system structured spaces can be both surprising and informative, the representations are characteristically derived from the data and are not designed to capture the user's thinking. They tend to be limited in terms of capabilities for recording the results of the user's thought processes. Also, displaying trends often necessitates a data reduction which can be at the expense of a rich representation of underlying data objects; this is in comparison, for example, with document surrogates typically used in search interfaces (Hearst, 2009). User-structured spaces use different means for recording the results of the users thinking and allow the user to reflect on that thinking, but they are typically limited in terms of their support for querying data and automatically structuring the results. INVISQUE explores the integration of user- and system-structured spaces within a single system.

4 Sensemaking

In the current study we report a qualitative analysis in which we drew on two models of sensemaking to help structure our observations. Both are descriptive models rather than prescriptive or normative models - and so it is not our intention to use them as a measure against which a system could be evaluated, but rather that they provided a conceptual vocabulary through which we interpret and articulate our results (where this was appropriately supported by the data). They are Pirolli and Card's model of intelligence analysis (Pirolli & Card, 2005), and Klein et al's Data/Frame model (Klein et al., 2006), both of which are described briefly below.

Pirolli and Card's model of intelligence analysis is based on interviews and think-aloud protocol analyses conducted with intelligence analysts. It depicts a workflow of component processes with two major loops: a foraging loop concerned with searching for and extracting information possibly into some kind of schema, and a sensemaking loop concerned with developing a conceptualization from the schema that best fits the evidence.

Significantly, the model is not committed to a single direction of processing but has multiple loops which move both from the bottom-up (data to theory) and from the top-down (theory to data) in an opportunistic interplay (a common theme in accounts of sensemaking, see for example Weick's account of sensemaking in organizations (Weick, 1995), Russell et al's learning loop complex (Russell, Stefik, Pirolli, & Card, 1993), Attfield and Blandford's model of sensemaking in corporate investigations (Attfield & Blandford, 2011) and Klein et al's Data/Frame model (Klein et al., 2006) described below).

From bottom to top, the analyst searches and sets aside relevant documents in a 'shoebox', nuggets are extracted from these and re-represented schematically in some form ('schematization'). A theory is developed and is ultimately presented to some audience. From top to bottom, new theories suggest hypotheses to be considered, the schemas are re-considered in this light, collected evidence is re-examined, new information extracted, and new raw data is sought.

Whereas Pirolli and Card's model describes a sensemaking workflow, Klein et al's Data/Frame theory gives an account of cognitive processes underpinning sensemaking. It is presented as a theory of how understanding (and also misunderstanding) begins, evolves and changes. Drawing on schema theory, the model distinguishes two kinds of entity involved in sensemaking: data and frames. Data are aspects of the world that a sensemaker experiences. A frame is a representation which offers an explanatory interpretation of that data. Frames may take many forms, such as stories, maps or plans.

The Data/Frame model proposes seven sensemaking processes:

- **Connecting Data to Frame** An initial frame is formed in response to new information (as a pre-conscious process).
- **Elaborating the Frame** Gaps in a frame are filled as more data is encountered.
- **Questioning the Frame** Unexpected data violates expectations created by a frame.
- **Preserving the Frame** Data which does not match a frame is dealt with by explaining it away or simply ignoring it.
- **Re-framing** Discrepant cues may suggest a new frame.
- **Comparing Frames** Multiple, mutually inconsistent frames are compared. What Feltovich et al. (Feltovich, Johnson, Moller, & Swanson, 1984) termed a logical competitor set (e.g. alternative medical diagnoses).
- **Seeking a Frame** The sensemaker cannot construct a frame (as a conscious process).

According to the model, as we encounter a situation a few key elements, or 'anchors', invoke a plausible frame in an inference which is characteristically abductive. By extending further than the observed data, a frame offers an economy on the data required for understanding, but also sets up expectations for further data that might be available. Exploration guided by the frame can elaborate it or challenge its expectations through the discovery of inconsistent data. Like Starbuck and Milliken (Starbuck & Milliken, 1988), Klein et al. argue that a frame acts as an information filter, not only determining what information is but what aspects of a situation will subsequently be noticed (i.e. what counts as data).

Both the Pirolli and Card model and the Data/Frame model describe processes of sensemaking. If there are differences then we see these as concerned more with how they generalise and what they emphasise rather than substantive differences. The Pirolli and Card model is a process model of sensemaking specific to intelligence analysis and which draws attention to external, behavioural aspects of the task as much as to internal cognitive aspects. It discusses the use of external data sources, and various collections and external representations that the user creates and manipulates as well as internal schemas and hypotheses. The Data/Frame model is more generic and focuses in more detail on cognitive operations. Nevertheless, both present the bottom-up and top-down structuring of information as central to sensemaking and insight. We return to the Pirolli and Card model and the Klein et al. model in the interpretation of our findings.

5 INVISQUE

INVISQUE stands for **IN**teractive **VI**sual **S**earch and **QU**ery **E**nvironment (described in more detail in (Stelmaszewska, Wong, Attfield, & Chen, 2010; Wong, Chen, Kodagoda, Rooney, & Xu, 2011; Kodagoda, Wong, Rooney, & Khan, 2012)). The system combines searching, automatic clustering and free-form manual manipulation of index cards on an infinite canvas (see Figure 1). Like the work by Strobelt et al. (Strobelt et al., 2009), the index cards act as visual surrogates (Kerne, Smith, Koh, Choi, & Graeber, 2008) for information objects such as documents, videos and images and each displays corresponding meta-data.

The combination of visual forms and interactivity is designed to facilitate flexible and rapid document triage and associated ad hoc analysis. INVISQUE has an Adobe Flash user-interface which retrieves data from a Java-based middleware through a JSP hosted on an Apache Tomcat server. This allows for a modular separation between content and presentation. In the following we describe the features of INVISQUE, which in some cases were specific to the current study.

5.1 Search Clusters

At the start of a new session, INVISQUE presents the user with an empty canvas. Double-clicking anywhere invokes a local search box which can be used to enter and submit a keyword query. The results are then presented as a visual cluster (search cluster) of index cards (see Figure 2).

Figure 1: A screen shot of *INVISQUE*. In this example index cards represent news articles. The central grouping is a user-defined cluster relating to a bioterrorism threat. Six surrounding clusters are search results (four are minimised). ‘My Shoebox’ is another user-defined cluster.

Figure 2: A search cluster of index cards retrieved from a keyword search for *vastopolis*. The six results considered the most significant are displayed as index cards.

In a search cluster the six most relevant articles are shown initially. For the VAST 2011 dataset of news articles, relevance to the query is determined using a term frequency/inverse document frequency calculation in which expected frequencies are expressed as log-likelihood (with a 6.63 significance threshold). Cards are ordered on both the X and Y axes according to user-configurable parameters. Ordering can be toggled from descending to ascending. The default axes used in this study were: relevance (descending left to right), and publication date (descending top to bottom). For this study, the meta-data displayed on each index card were: expected relevance (log likelihood of keyword), title, date of publication, the top three keywords (based on log likelihood calculation), physical locations mentioned in the text, and the keyword that was matched (prior to stemming).

Clicking a document icon at the top right of an index card opens the underlying data object, whether this be text, image or video, as an overlay on the canvas.

We show the six most relevant index cards initially to reduce visual clutter. An overview of the complete result set is shown on a contextual interval slider bar associated with each search cluster on which all results are shown as points (see Figure 2 for the slider bar in context and enlarged in Figure 3). Points are organised by the same attributes as the index cards. Users can use the slider to adjust a 'window' into the data, changing the index cards that are displayed. They can also scroll through the dataset, click on a specific location in the dataset, or increase/decrease the width of the window (changing the number of visible index cards).

In addition to typing keywords, a new search can be conducted by grabbing an item of meta-data from an index card and releasing it onto the canvas. This performs a fresh search on the dataset and produces a new cluster at the position where the meta-data was released.

The canvas allows users to create and work with multiple search clusters with the freedom to organise the layout by dragging. Opacity and transparency differentiate the cluster which currently has the focus. Clutter can be further controlled by minimizing clusters to show just the title and interval slider bar.

Figure 3: The contextual interval slider.

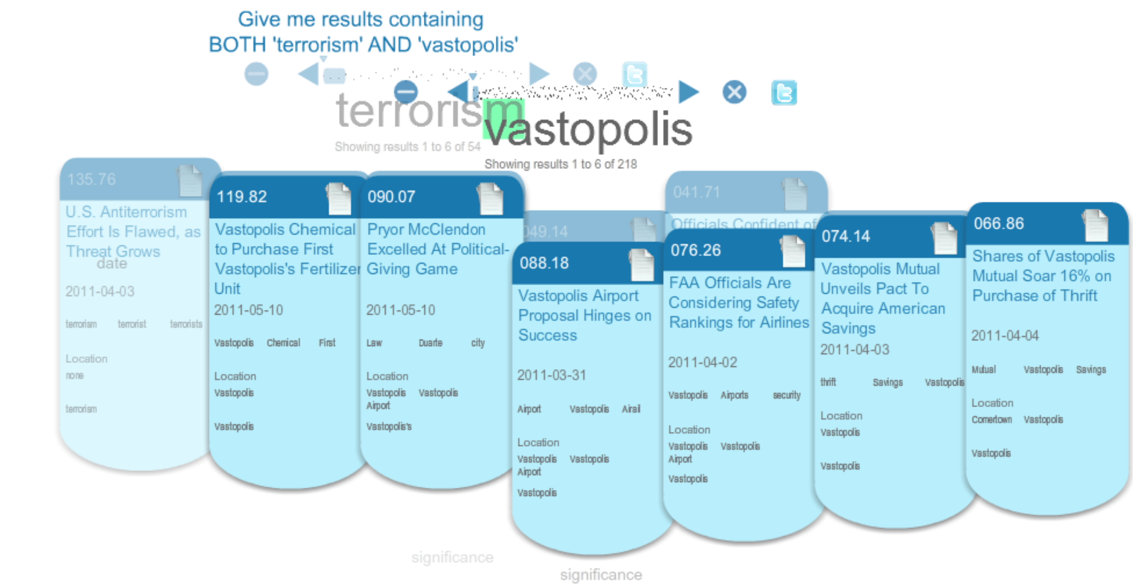
5.2 Combining Searches with Boolean Operations

Boolean logic is often used with faceted search to widen or narrow the search window (Seifert, 2011), but explicit Boolean operations can be hard to understand. One solution to this is to make Boolean operations implicit in the faceting process (Jetter, Gerken, Zöllner, Reiterer, & Milic-Frayling, 2011). INVISQUE allows the user to create Boolean combinations of searches implicitly by the spatial manipulation of search clusters.

For example, two results clusters can be combined to show either their intersection (AND), or union (OR) by dragging the title of one cluster over another. Releasing a cluster when there is a small overlap between cluster titles performs an intersection (see Figure 4(a)). Releasing the mouse when there is a large overlap represents a union (see Figure 4(b)). These operations use a Venn diagram metaphor, where a small overlap represents the intersection, and a large overlap represents the union.

Both operations create a new cluster. As an aid to understanding, the expected outcome (dependant on the overlap) is described above in natural language while one cluster is held over another. Colour and a bounding rectangle are used to further clarify the difference.

NOT operations on meta-data can also be performed by grabbing the meta-data on the card with the mouse cursor and shaking it (as if to say ‘no’). This has the effect of removing all cards from the cluster that contain the same meta-data.



(a) Performing an *AND* operation.

(b) Performing an *OR* operation.

Figure 4: Performing Boolean operations by dragging one cluster title over the other.

5.3 Identifying Relationships

A user can identify index cards with a common meta-data property by clicking a meta-data value on a single card. When this is done all index cards that do not share this value are faded to 50% opacity, and occurrences of the value are visually highlighted in red. Meta-data highlighting shows relationships both within clusters, and across all clusters on the canvas (see Figure 1 for an example). Objects on the contextual interval slider that share the property are also highlighted (see Figure 3), allowing relationships to be seen with hidden cards and cards in minimised clusters.

5.4 User-defined Clusters

Index cards make their first appearance on the canvas as part of a search cluster. However, cards can be selected individually and dragged to form user-defined clusters. As an index card is moved from its native search cluster, a 'wobble' animation indicates a breaking of the cluster metaphor. The index card can then be placed anywhere on the canvas, and other cards added to this new collection.

Once the groups of cards reach a certain size, they themselves may need to be grouped and sorted (see Figure 5). By drawing a line through the cards the user indicates to the system that they are to be treated as a new cluster allowing automated ordering functions to be selected and applied (e.g. date of publication). These user-defined clusters can then be renamed.

A bespoke cluster, of course, can be used to map to any concept that a user finds convenient. Possibilities include a set of potential leads, a collection of primary evidence, secondary evidence, or counter evidence to an evolving theory - whatever is useful within an ad hoc, evolving and reflexive investigation process. Importantly for the sensemaking process, user-defined clusters could be used to map to the semantics extending beyond the documents per se to include hunches that become the focus of a line of investigation.

Information visualisation tools typically support automated structuring, whereas mind-mapping and argument mapping tools are typically user-led. Part of the motivation for INVISQUE was to allow the user to move opportunistically and seamlessly between both modes of interaction. The aim is to lend lightweight, automated support to the organisation (or schematization (Russell et al., 1993)) of manually generated collections of data so that patterns and gaps can be explored.

5.5 Similarities to Other Systems

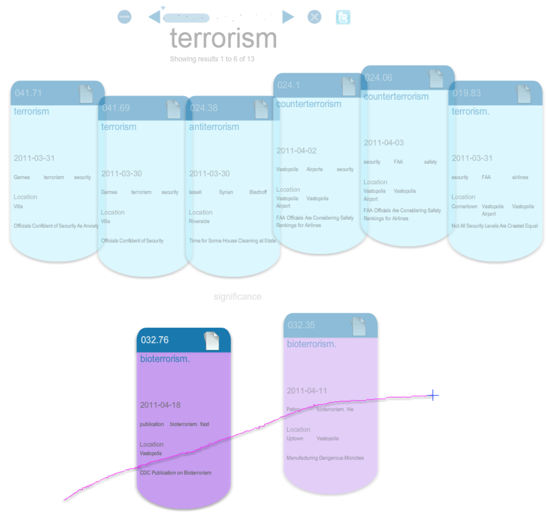
We see INVISQUE as being similar to the Analyst's Workspace developed by Andrews et al. (Andrews & North, 2012). The systems differ in that INVISQUE is designed for smaller displays by allowing the user to pan and zoom around the infinite canvas, and by presenting documents as surrogates to save on display real estate. Also, INVISQUE has a greater focus on the system-structured clustering where documents can be ordered in both the X and Y axes, and clusters can interact with one another - resulting in Boolean operations.

6 User Study

We conducted an observational study to investigate how domain experts perform a sensemaking task using the INVISQUE system. The aim of the study was two-fold: (i) to observe and understand the strategies that experts might use to forage through, make sense of, and hypothesise about data, and (ii) to observe how the functionality provided by INVISQUE supports/hinders them through this process. We chose experts in intelligence analysis because, in the interests of ecological validity, we believed it would provide greater insight into how sensemaking is performed in the real world, thus providing us with a better understanding of the role tools such as INVISQUE can play.

Similar studies have been performed with the Jigsaw (Kang, Gorg, & Stasko, 2009) and Force-SPIRE (Endert, Fiaux, & North, 2012a) systems. In comparison, our study presents participants with a text corpus of over 4000 articles (as opposed to 50 and 111 respectively). A larger collection means that it is not feasible to read each document individually, and so a different investigative approach may be required.

(a) Two index cards are extracted from the cluster.



(b) *Connecting the dots* is used to generate a new a new cluster.

(c) The new cluster is created.

(d) The name of the cluster is redefined by the user.

Figure 5: Collating index cards into primary evidence.

6.1 Method

6.1.1 Participants

Six participants with experience in intelligence analysis participated in the study. They were male, English speakers. The study was carried out in two separate sessions with three participants in each. The first group (Group A) comprised members of a government team who provided capabilities development for intelligence analysts - hence they were alert to the nature of intelligence analysis tasks and its system requirements. The second (Group B) were defence science researchers with experience of intelligence analysis from a different government agency.

6.1.2 Materials

Analysts interacted solely with the INVISQUE system during the study using mouse and keyboard. The studies were carried out on-site, and the host organisations provided the displays. The first group used a 20” monitor with a resolution of 1600x1200 pixels (1.9 million pixels of display real estate), while the second group used a 40” monitor with a resolution of 1600x900 pixels (1.4 million pixels of display real estate).

The evaluation was designed around the Mini-Challenge three dataset of the 2011 IEEE VAST challenge (Maryland, 2011). The data set consists of 4474 fictitious news articles. The data contained a known ground truth as well as several garden paths. None of the participants had a previous exposure to the dataset.

6.1.3 Procedure

Although we were not interested in studying collaboration, we intentionally structured sessions as group co-discovery sessions (three participants per group) (Dumas & Redish, 1999). This was to encourage think-aloud and general discussion (rationale for actions etc.), which in turn helped us map thinking processes to actions at the interface. Also, time with the participants was limited, so working collaboratively maximised the amount of time that each participant could spend on the task. The study setup was not designed in a way that was sensitive to how collaborative intelligence analysis is performed in practice, and so we saw no benefit from analysing collaborative issues.

Sessions were carried out in two parts. In the first, participants were instructed on how to use the system, making use of one of the dead-end paths in the dataset as an example walk-through. All interactions were performed by the participants, to encourage learning by practice.

Immediately after the 20 minute practice session, the canvas was cleared and participants were told that a flu outbreak had occurred in the city of Vastopolis and that the authorities suspected that it may have been a result of terrorist activity. Participants were asked to use the system to determine a cause for the outbreak. They were also asked to think-aloud as they reasoned through the problem and to explain what actions they did and why. Due to time constraints, if the participants went completely off track, the session was interrupted, and they were given clues to put them back on track.

6.1.4 Data Collection and Analysis

For reasons of security, we were not able to record video or audio during the sessions. We were, however, able to perform full-resolution screen-capture recordings, and two researchers recorded observational notes at each session.

We processed the data in three stages. Firstly, all the user interactions that occurred were logged and time stamped. Secondly, we used these logs in combination with the the observational notes to provide explanations of the participants' interactions and reasoning. Finally, we visualised the actions performed as an activity time line (see Figure 6), which we explain in detail below.

The data were then analysed by coding for emergent themes. Codes were used which mapped to constructs within the Pirolli and Card (Pirolli & Card, 2005) and Klein et al. (Klein et al., 2006) models where they offered a valid interpretation of part of the data. The approach was inductive and interpretive. Differences and similarities were extracted from both observational sessions, and led to five core themes, which form the sub-headings of the Findings section. The first three relate to activities in Pirolli and Card's model of sensemaking (Pirolli & Card, 2005), the latter two are more general observations of participant's behaviour.

6.2 Visualising Activity

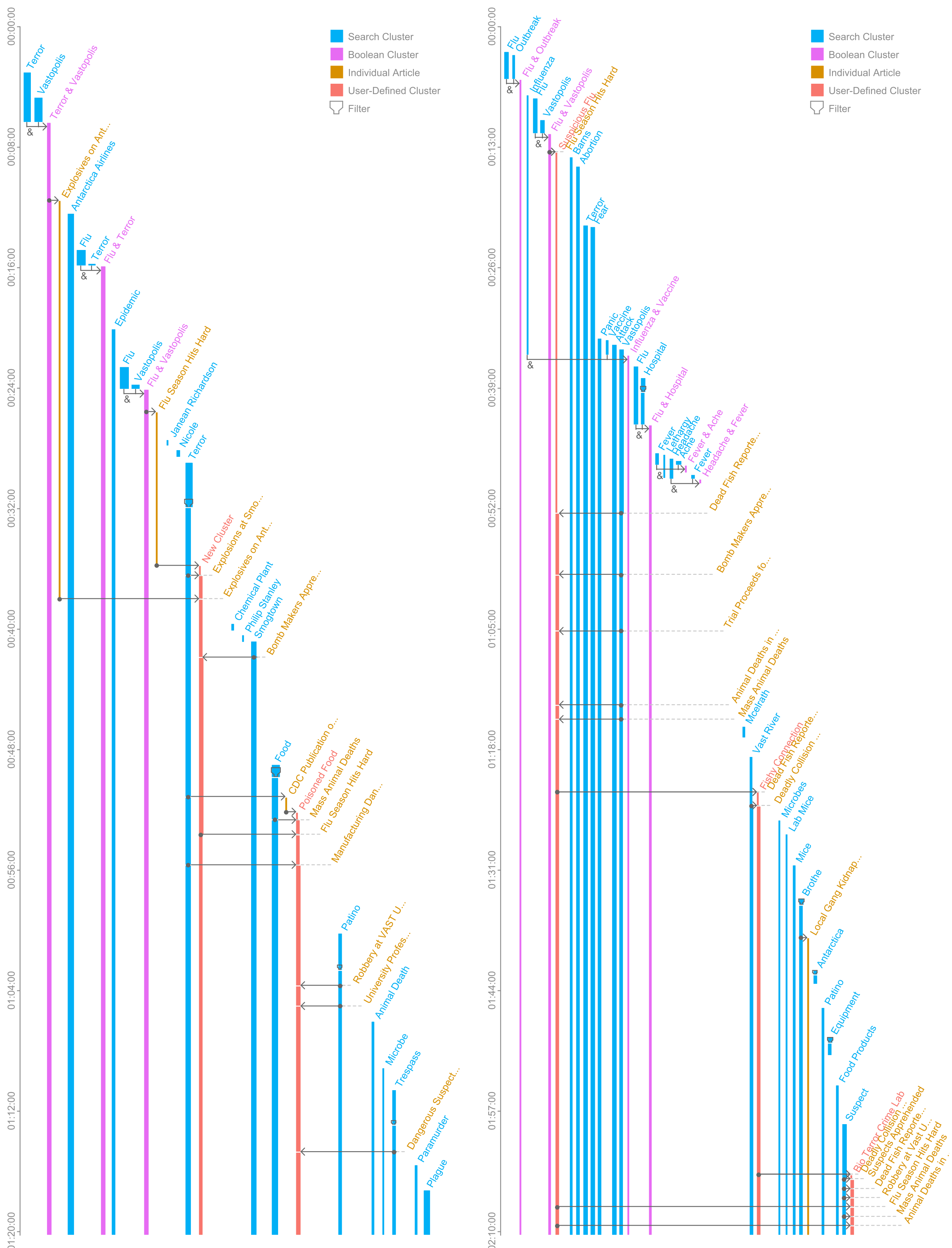
Figures 6(a) and 6(b) show the flow of activity that Groups A and B performed respectively. They show when a new search was performed (blue bars); when a cluster was filtered (funnel icon above a reduction in bar width) or a Boolean operation was performed on two clusters (arrow to a pink bar); when individual articles were extracted from a cluster (arrow to a brown bar), or moved from one cluster to another (arrow to a blue or red bar); and when user-defined clusters were created (red bars). The figures show progression from top-to-bottom, with each new activity placed in an adjacent column. Each bar represents a cluster, with height representing the period of time the cluster existed on the canvas and the width representing the number of articles in the cluster. Due to a large variation in the number of articles contained in the clusters, we used a non-linear mapping between the number of articles and the width of the bar. The labels represent either a search term, Boolean operation, title of an individual article, or the name of a user defined cluster.

6.3 Findings

In the following, we discuss four kinds of activity we observed: *Search*, *Constructing Shoeboxes*; *Constructing Explanatory Narratives* and *Date Ordering*. We additionally discuss observations of *Visual Persistence*.

6.3.1 Search

Figure 6 indicates that both groups performed keyword searches throughout the sessions. We note that during both sessions the searches changed in character as the sessions progressed. Figures 7(a) and 7(b) illustrate this by showing searches laid out consecutively and by the number of results returned. Although there is some fluctuation, there is also a general trend for searches during the early stages (e.g., Terror:105, Vastopolis:218, and Flu:338) to return more results than searches during the later stages (e.g., Microbe: 2, Food Products:7, and Patino:5). This, and the content of



(a) Group A

(b) Group B

Figure 6: Activity time line for both groups, with time ascending top-to-bottom on the Y axis. Bar width represents the number of articles contained within cluster.

the search terms, indicates a trend for information needs to become more narrow and better defined as the task progressed.

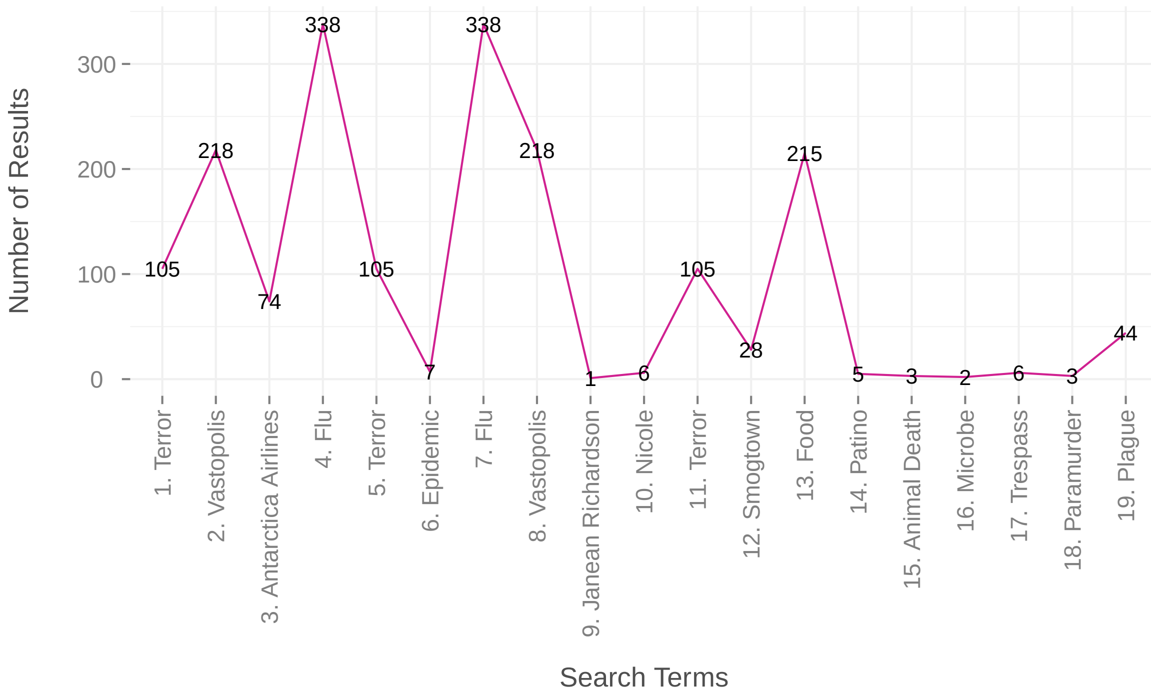
Progression towards more narrowly specified information needs corresponded to groups focusing on the development of specific explanatory narratives, each corresponding to the interpretation of a sequences of events over time (see Section 6.3.3 below on explanatory narratives). During the early stages, participants often used intersection operations (Boolean AND) to combine and narrow existing search clusters. For Group A, the ability to do this led to the discovery of a group of articles that they found to be of some interest (see ˜12 minutes and ˜25 minutes in Figure 6(a), where articles on an airline explosion and a flu outbreak were discovered).

Notably, during their respective sessions, each group settled on a single search cluster from which they drew a good deal of the content, which they would then focus on later in relation to a number of enquiry streams. For Group A, this was a search for ‘Terror’, from which they extracted four articles, and for Group B it was a search for ‘Vastopolis’, from which they extracted five articles. Both groups returned to these clusters repeatedly over the course of the investigation and also used these keywords a number of times throughout. Even when using different search terms, both groups were still able to identify key articles that related to the ground truth, although this may be considered an artefact of the data rather than the system.

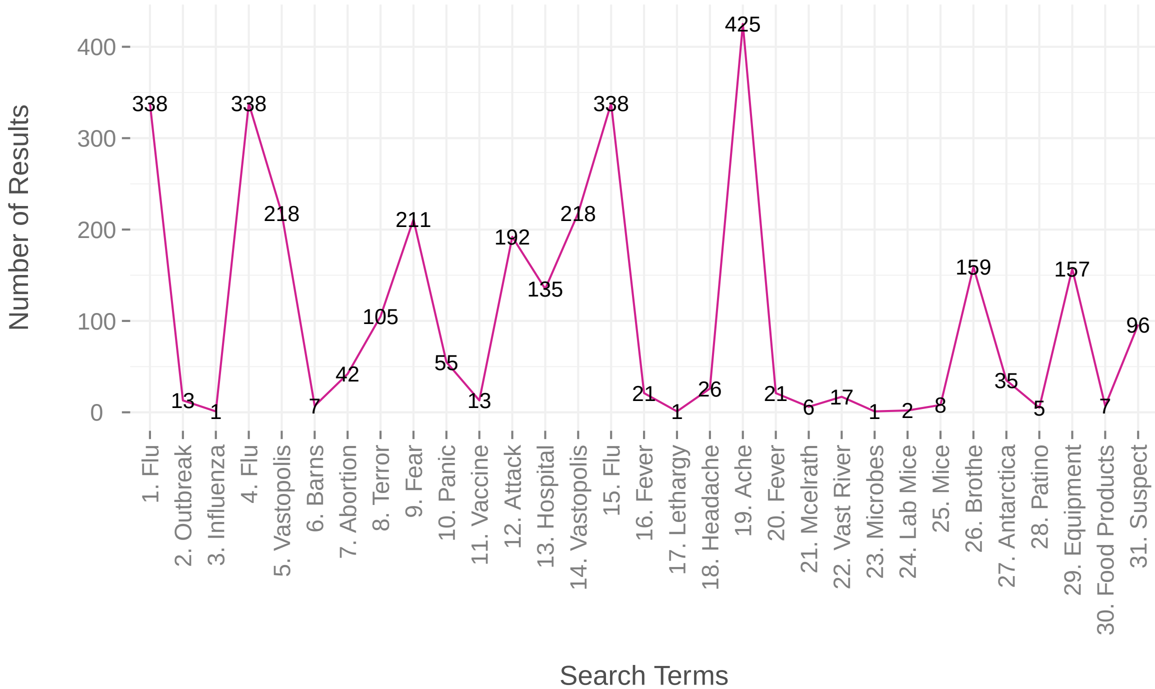
6.3.2 Constructing Shoeboxes

As each group came across articles in search clusters that interested them, they separated them out spatially into new, user-defined clusters. Interesting articles were ones that allowed the group to construct some kind of theory about terrorist activity, although early on these theories were characteristically ill-defined and difficult to articulate. Each group established and explored multiple lines of investigation, or theories.

We note two distinct ways in which the interface canvas was appropriated for structuring document triage and subsequent investigation corresponding to depth-first and breadth-first strategies. Group A adopted a depth-first strategy. When they came across an article of interest, they set it



(a) Group A



(b) Group B

Figure 7: The search terms used by both groups (ordered chronologically) and the respective number of results.

aside and immediately performed searches for reports which might elaborate an explanatory theory that the initial article had provoked (i.e. elaborating the frame). Where little or no supporting evidence was found, or evidence was found that was inconsistent with the theory, it was dropped, closing off that line of inquiry. When this happened, the group returned to exploring documents in a prior search to consider new possibilities. This is illustrated in Figure 6(a). Between 8 and 40 minutes, articles describing an explosion on an aircraft, a flu outbreak and an explosion at Smogtown chemical plant were extracted from search clusters. Each extraction is followed by related searches, until the lines of enquiry were dropped in favour of something new.

In contrast, Group B used a breadth-first strategy. This involved systematically triaging search results in one go, setting aside all documents that might be useful in a single collection. They referred to this collection as a 'shoebox' (having some familiarity with Pirolli and Card (Pirolli & Card, 2005)). In Figure 6(b) the creation of the shoebox can be seen at minute 13. The group spent approximately 20 minutes scrolling through the Vastopolis search cluster, moving potentially relevant articles into the 'Suspicious Flu' cluster. At this stage, no connections between articles were made. Once the initial 'triage' or 'rough cut' was complete, they then considered each shoebox article in turn, performing further searches to find related information.

In terms of the Klein et al. model (Klein et al., 2006) our interpretation of both strategies is that articles cued or in some way contributed to a frame worthy of elaboration. The participants worked through a range of possible, and potentially competing frames (in sequence or in parallel), giving shape to each in the process. In both strategies, participants used the 'setting aside' behaviour to reify their frames and, in the case of the depth first strategy, as markers or reminders to elaborate them further. The value of the interface was to provide a low-cost method for managing the complexity of systematic triage and multiple areas to follow-up.

6.3.3 Constructing Explanatory Narratives

Following triage, the groups organised information into clusters of related documents organised in chronological order (i.e. schematisation (Pirolli & Card, 2005)). Documents represented within each cluster were related insofar as each provided information that contributed to a single explanatory narrative.

An example of this can be seen around minute 52 (see Figure 6(a)) when Group A established the Poisoned Food cluster. They then continued to add other articles that may have been related through the emerging narrative. Figure 8 shows a screenshot of this cluster. Around it lies a residue of search clusters that provided its raw material. The central cluster contains seven articles. These report: (i) A significant flu outbreak in the city of Vastopolis; (ii) A university professor walking out of class due to stress; (iii) An arrest at a local food processing plant; (iv) A robbery at the university where equipment was stolen from the Professors office; (v) The ease with which one could carry out a terrorist attack; (vi) A lecture on the ease of manufacturing dangerous microbes; and (vii) Mass animal deaths at a local food processing plant. Participants within this group articulated an associated explanatory narrative which we paraphrase as:

(a) Terrorists stole equipment from a university laboratory. (b) They used this to develop dangerous microbes. (c) They released the microbes into the local food chain causing the outbreak.

Notably, the narrative extends beyond the information presented in the articles, but is (apparently) consistent with it. For the most part, it offers a framework for explaining the information found. For example, proposition (a) is a potential explanation for the robbery report, and proposition (c) is a potential explanation for the outbreak, arrest and mass animal deaths. Parts of the narrative also appear to be explained by the document. For example, proposition (b) is explained (partially) by the ease of manufacturing dangerous microbes, and proposition (c) is explained (partially) by the ease of carrying out a terrorist attack. Hence we note that, although the document cluster is closely associated with the narrative, the cluster and the narrative are not the same but stand in a relationship of apparent coherence; there is a set of articles chronologically arranged, and then there is a narrative which both explains and is explained by information within those articles.

As new articles were discovered, offering new information concerning any given narrative, that

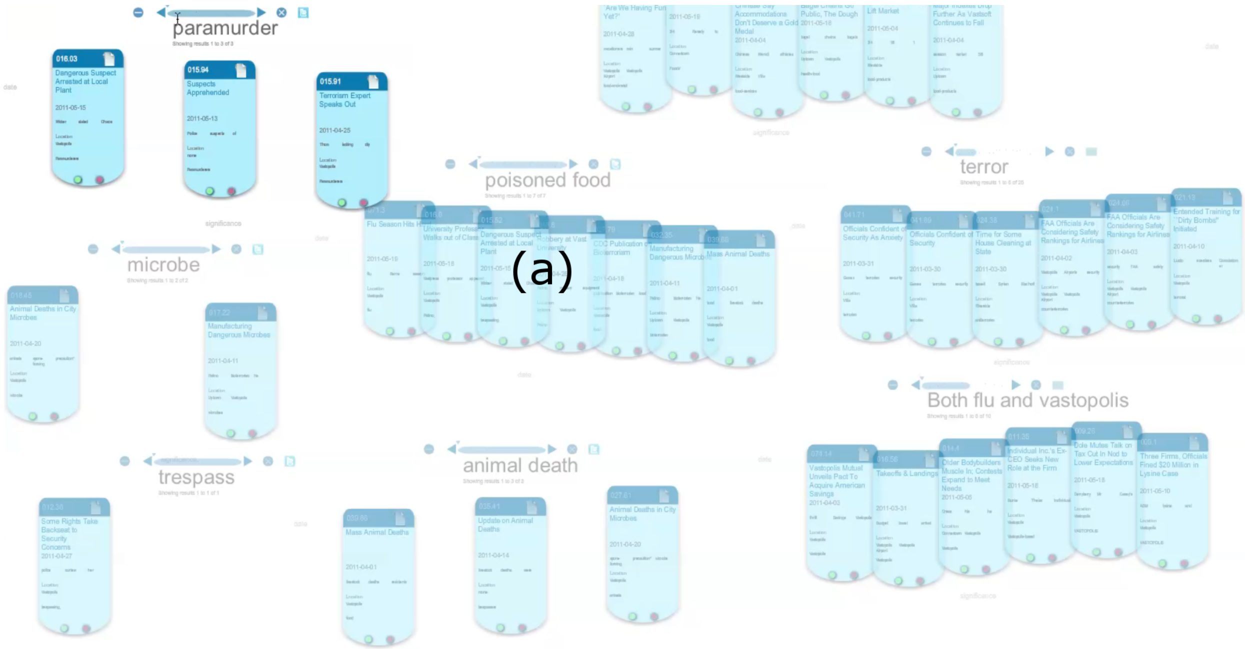


Figure 8: A screen capture showing the use of space during the first session. The analysts are currently searching for articles on paramurder to elaborate on their current line of inquiry (a), which is in the middle of the canvas.

narrative evolved in the light of the new data (i.e. elaborating the frame). And in some cases information was found that led the groups to question a narrative (i.e. questioning the frame). For example, we observed Group B reach a point when their explanatory narrative (Fishy Connection) no longer appeared to them to be the best explanation for the data, and a new narrative was offered in its place. To symbolise this, a new cluster was created (Bio Terror Crime Lab) and articles from the previous narrative, recent searches and the original shoebox (suspicious flu) were reorganised to create a new chronology (see Figure 6(b) at ˜2:00 hours).

6.3.4 Using Date Ordering

When a search is conducted, INVISQUE returns articles as a cluster with index cards ordered on the X-axis by an automatic relevance calculation and on the Y-axis by publication date. Ordering on either X- or Y-axis can then be changed by the user. We observed during the study that participants almost exclusively focussed on and exploited temporal ordering. Group A re-ordered clusters by date on the X-axis nine times and Group B did this ten times. A participant from Group B commented that temporal ordering was a "force of habit", with x-axis date ordering applied to both search clusters and user-defined clusters.

Ordering documents in this way supported participants in reviewing documents in terms of a posited explanatory narrative. With the cards ordered by date of publication, they accessed the full text of each in turn and chronologically, reading the document aloud and discussing the part of the narrative that had been associated with it. These 'readings' prompted group discussion about the narrative's credibility by drawing attention to questions of coherence (questioning the frame). Reviewing documents in this way also supported the identification of gaps in the narrative and expectations that it might have generated, leading to new questions and searches that might help to elaborate or test the narrative (elaborating a frame).

Date-ordering also appeared to assist participants in testing causal hypotheses as part of testing narrative/information coherence. For example, Group A developed a theory that a chemical plant explosion had contributed to spreading pathogens. They did this based on an article reporting the explosion and another reporting the epidemic. As a test, they then put the reports into a single cluster and sorted them by date. This offered them a falsification strategy, based on the premise that a causal relationship implies a specific temporal order. Epidemic reports followed by the explosion would have been inconsistent with the proposed narrative. As it happened, the result was consistent with the idea of the explosion having a causal role and the theory was maintained (this can be seen at ˜35 minutes in Figure 6(a)).

Interestingly, the system's default date ordering was descending from left to right. This did not appear to bother the analysts. They had the facility to swap the ordering, but did not take advantage of it. Descending temporal ordering can be seen in the Poisoned Food user-defined cluster in Figure 8.

6.3.5 Retracing Steps in the Visual Record

Visual persistence of clusters on the interface canvas enabled participants to keep track of their searches over time. Periodically they went back to review query terms they had used, the queries themselves, and the relevance of the search results. The bars in Figure 6 indicate that many clusters remained on the interface canvas until the end of the session (although they were not always in view due to panning and zooming).

Retaining multiple searches on a single canvas also allowed participants to return to articles they remembered seeing but did not have a use for previously. We return to the data/frame theory to interpret the value of this. According to that theory, as a frame develops, questions emerge. Previously seen information may well be relevant to these and may need to be re-found. This was observed a number of times. For example, after spending time searching for, and examining, articles relevant to the chemical plant explosion theory, and at the point that they were attempting to find an explanation for this consistent with an act of terrorism, the group who worked on this theory (Group A) recalled an article they had seen earlier that reported traces of explosives being found in the wreckage of an air crash. By visually scanning across the canvas they were able to re-find this article quickly and move it to the explosion theory cluster they were developing (this can be seen at ˜38 minutes in Figure 6(a)). In fact, the action of extracting articles from prior clusters can be seen throughout Figure 6, highlighting the value of maintaining visual persistence.

Participants also commented on the way that the free-form visual organization of the canvas offered a visual record of their thinking. In particular, they spoke of the traces of searches and articles across the canvas providing a visual reminder of dead-ends or 'rabbit holes' (i.e. search strategies and explanatory narratives that had yielded little positive or contradictory evidence). By implication this also helped them maintain a visual record of fruitful pathways. A similar observation as made by Andrews et at. who found that the spatial layout of articles supported recall and allowed users to generate informal relationships (also known as semantic layering) (Andrews et al., 2010).

Figure 8 shows how Group A placed a cluster in the centre of the canvas (Poisoned Food), and used the surrounding space to further investigate an associated narrative by performing keyword searches. Here the cluster offers a visual anchor, drawing in relevant articles from the surrounding clusters, which form (almost accidentally) a radial layout surrounding the main cluster.

6.4 Discussion

The user-study we have reported was exploratory, but provides encouraging evidence for the value of an interface that integrates search with free- and system-supported manipulation of information objects for supporting investigatory sensemaking tasks such as those performed as part of intelligence analysis tasks. The combination of flexible manipulation, ad hoc structuring and visual persistence enable participants to follow, evaluate, drop and develop multiple lines of enquiry.

We were able to interpret user-actions and related cognition in terms of concepts from both the Pirolli and Card (Pirolli & Card, 2005) and the Data/Frame (Klein et al., 2006) models. For the former, we observed periods of foraging, where participants search, filtered and populated their 'shoebox'. As the investigation progressed, they moved into sensemaking phases of schematisation and hypothesis generation orientated around the construction of explanatory narratives. Movements between these activities appeared to be natural and seamless. It appeared that the free-from canvas of INVISQUE, and the interplay between search and user-defined clusters, supported this behaviour and allowed for fluid movements from foraging to sensemaking and back again.

Centre to our findings was the creation and use of user-defined clusters. Participants appropriated these as 'shoeboxes', enabling systematic, yet flexible, triage actions in ways that organised information for deeper analysis. They also used clusters as timelines to support their thinking about documents in terms of explanatory narratives. These narratives, or stories, which both explained and were explained by the documents can be thought of as frames (Klein et al., 2006). Optional date ordering allowed participants to systematically review, elaborate and test these frames.

Except for moving clusters around to create space, the participants were seldom required to interrupt their investigation, or work around the system, while organising the data. Once organised, they were able to review and question the frame they created, and to use it as a common visual artefact to discuss amongst themselves and to point to certain aspects or a point in the analysis to question the assumptions or data used. This rapidly allowed them to question the frame, or to preserve parts of the frame while elaborating on it.

The visual representation apparently helped participants externalize and make tangible what is normally hidden and transparent within the thinking process. Visual persistence on the system canvas provided a record of thinking over time and freed the analysts from having to remember what had already been considered. The system appeared to support ‘recognition over recall’ of the thinking process itself.

From a usability perspective, a number of changes could be made to improve the system. After performing Boolean operations, participants in both groups were often required to recreate clusters after the system replaced both search clusters with the intersection result. This prevented them from reusing the component searches in new Boolean combinations (something they wanted to do), thus increasing their workload. Also, removing the clusters without prompt breaks the visual record. Rather, the system should create a third cluster as the result of a Boolean operation, allowing the original clusters to persist on the canvas.

The discovery of some articles also led to searches for specific people or locations. Automated entity extraction (as seen in tools such as the Analyst’s Workspace (Andrews & North, 2012)) allows users to identify the people and places in an article, and further investigate their occurrence in the dataset. However, this was not built into the INVISQUE system and is something that participants had to perform manually by typing the names into new keyword searches. An example of where manual entity extraction failed can be seen at ˜13 minutes in Figure 6(b), where *Nicole Barns* was identified in the *Flu Season Hits Hard* article. This led to a search for *Barns*, which in turn led to the discovery of an article describing an abortion drug, leading to a keyword search for *abortion*. The key problem with this line of enquiry being that a Dr. Lynne Barns was mentioned in the abortion article (not Nicole), meaning that the participants established a fictional relationship between flu and abortion.

Date ordering was constantly used by our participants, but the INVISQUE system orders index cards equidistantly, such that a date ordering is essentially a chronology. This makes it difficult to identify gaps in the timeline. Imposing a true temporal scale, such that articles published at a similar time are placed spatially close, may have helped to highlight where there is missing data in the narrative.

7 Implications for Design

Based on our observing of how participants appropriated features of INVISQUE in the service of the investigatory sensemaking task that we studied, we draw a set of more generalisable conclusions about features that users are likely to find helpful for this kind of task. Each could be considered as hypotheses for future evaluation. They are:

Seamless Interaction - The ability to move seamlessly between foraging and narrative construction (or schematisation in general) within a single workspace appears valuable, allowing for fluid transition between the pursuit of hunches and evolving explanations. Our participants constructed fairly systematic approaches to the investigation, and yet their movements between foraging and sensemaking appeared relatively unencumbered by the interface.

Ability to impose structure - Whilst we believe that free-form exploration and arrangement of information is crucial in sensemaking, it was clear that our participants relied heavily on the clustering of articles and the imposition of system-structured space. It was the ability to enforce chronological ordering that allowed them to cue, elaborate and evaluate explanatory narratives.

Depth first vs. breadth first - Interface design typically places implicit constraints on task structure. However, free-form organisation at the interface appears to provide for flexibility, such that users can create systematic task structures on the fly as it suits them. For sensemaking tasks, this can include the ability to triage documents and pursue resulting lines of enquiry using both depth-first and breadth-first strategies as required. Whether one strategy is better than the other is perhaps an open question. Nevertheless, the flexibility to adopt one or the other based on task constraints or user preference appears valuable.

The Importance of Time - From our study, it is apparent that temporal ordering of information during investigatory sensemaking is important. Timeline views appear important for visualizing how events have evolved over time and construcing narratives which can explain and be explained by the known facts. For example, we have seen that timelines can assist in identifying/testing causal relationships, and identifying gaps in the data.

But whilst equidistant ordering may maximize the readability, it may make if difficult to identify the gaps in the sequence. Since this study we have added the functionality to toggle between equidistant and temporal positioning of index cards. We notice that temporal positioning affords an additional capability of identifying and evaluating potential causal relationships based on temporal separation.

Persistence and the visual record - Allowing an investigative task to take place on a large, or infinite, canvas provides a real estate where data can persist. The importance of spatial memory for sensemaking tasks has been identified by Andrews et al. (Andrews et al., 2010), and we see evidence of its importance in our study. As theories and narratives evolve over time, analysts are able to use this spatial memory to return to, and incorporate, important documents that were discovered earlier. Similarly, the visual record makes it easy to retrace steps once an avenue of investigation has reached an end, and a new line of enquiry is required.

8 Conclusion and Future Work

The implementation of INVISQUE and the study reported here represents one step in a task-artefact cycle in which we aim to evolve the INVISQUE design concept into a general purpose reasoning workspace for supporting a range of sensemaking problems. INVISQUE uses index cards grouped in clusters to represent data items, which when combined with the ability to be arranged freely, offers both system-structured and user-structured spaces. The former improves the process of sifting through raw data and analysing timelines of events, whilst the latter allows users to collate evidence for creating shoeboxes or proceeding down new lines of inquiry.

In conclusion, INVISQUE has been able to combine visual forms in a somewhat novel way that allows analysts freedom to organise data, their lines of inquiry, and to construct and review explanatory narratives. This freedom to use the space, in addition to the various underlying technologies, appears to be an important feature in supporting sensemaking in investigative domains.

Not all sensemaking tasks are the same. The task used in the current study was seeded by a hypothesis and this provided participants with obvious entry points into the data (e.g., keyword searches for *terror* or *vastopolis*). More exploratory tasks may benefit from greater overviews of the underlying collection. As future work, we are developing INVISQUE to provide a range of collection overviews, allowing for more data-driven (bottom-up) investigations. Different kinds of view may also be important for cueing and representing different kinds of frames. We are currently exploring the role that different views can have in exposing different features to the user and making more systematic comparisons between competing theories. These will provide a basis for future evaluation studies.

References

Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking: tight coupling of dynamic query filters with starfield displays. In *Chi* (p. 313-317). New York, NY, USA: ACM.

Andrews, C., Endert, A., & North, C. (2010). Space to think: large high-resolution displays for sensemaking. In *Chi* (p. 55-64). New York, NY, USA: ACM.

Andrews, C., & North, C. (2012). Analyst's workspace: An embodied sensemaking environment for large, high-resolution displays. In *Visual analytics science and technology (vast), 2012 ieee conference on* (p. 123-131).

Attfield, S., & Blandford, A. (2011). Making sense of digital footprints in team-based legal investigations: The acquisition of focus. *Human Computer Interaction*, *26*(1-2), 38-71.

Bauer, D., Fastrez, P., & Hollan, J. (2005). Spatial tools for managing personal information collections. In *Hicss* (p. 104-113). Washington, DC, USA: IEEE Computer Society.

Billman, D., & Bier, E. (2007). Medical sensemaking with entity workspace. In *Chi* (p. 229-232). New York, NY, USA: ACM.

Dumas, J., & Redish, J. (1999). *A practical guide to usability testing*. Intellect, Limited.

Endert, A., Fiaux, P., & North, C. (2012a). Semantic interaction for sensemaking: Inferring analytical reasoning for model steering. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(12), 2879-2888.

Endert, A., Fiaux, P., & North, C. (2012b). Semantic interaction for visual text analytics. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 473-482). New York, NY, USA: ACM.

Feltovich, P., Johnson, P., Moller, J., & Swanson, D. (1984). The role and development of medical knowledge in diagnostic expertise. *Readings in Medical Artificial Intelligence: The First Decade*, 275-319.

Gregory, M., Payne, D., McColgin, D., Cramer, N., & Love, D. (2007). Visual analysis of weblog content. In *Weblogs and social media* (p. 2-4).

Hearst, M. (2009). *Search user interfaces*. Cambridge University Press.

Heuer, R. (1999). Psychology of intelligence analysis. *Intelligence*, *284*(5416).

Jetter, H., Gerken, J., Zöllner, M., Reiterer, H., & Milic-Frayling, N. (2011). Materializing the query with facet-streams: a hybrid surface for collaborative search on tabletops. In *Chi* (p. 3013-3022). New York, NY, USA: ACM.

Kang, Y., Gorg, C., & Stasko, J. (2009). Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study. In *Visual analytics science and technology, 2009. vast 2009. ieee symposium on* (p. 139-146).

Kang, Y., Grg, C., & Stasko, J. (2010). How can visual analytics assist investigative analysis? design implications from an evaluation. *IEEE Transactions on Visualization and Computer Graphics*, *17*(5), 570-583.

Kerne, A., Smith, S., Koh, E., Choi, H., & Graeber, R. (2008). An experimental method for measuring the emergence of new ideas in information discovery. *International Journal of Human-Computer Interaction*, *24*(5), 460-477.

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2006). A Data/Frame Theory of Sense Making". In *Expertise out of context: proceedings of the sixth international conference on naturalistic decision making* (p. 113-155).

Kodagoda, N., Wong, B. W., Rooney, C., & Khan, N. (2012). Interactive visualization for low literacy users: from lessons learnt to design. In *Proceedings of the 2012 acm annual conference* *on human factors in computing systems* (pp. 1159–1168). New York, NY, USA: ACM.

Marshall, C., & Shipman, F. (1997). Spatial hypertext and the practice of information triage. *Hypertext*, 124-133.

Maryland, U. of. (2011). *IEEE VAST Challenge 2011.* `http://hcil.cs.umd.edu/localphp/hcil/vast11/`. (Online; accessed October 31st 2012)

Ntuen, C., Park, E., & Gwang-Myung, K. (2010). Designing an information visualization tool for sensemaking. *International Journal of Human-Computer Interaction*, *26*(2-3), 189-205.

Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Intelligence analysis.*

Russell, D., Slaney, M., Qu, Y., & Houston, M. (2006). Being literate with large document collections: Observational studies and cost structure tradeoffs. In *System sciences* (p. 55-62). Washington, DC, USA: IEEE Computer Society.

Russell, D., Stefik, M., Pirolli, P., & Card, S. (1993). The cost structure of sensemaking. In *Chi* (p. 269-276). New York, NY, USA: ACM.

Seifert, I. (2011). A pool of queries: Interactive multidimensional query visualization for information seeking in digital libraries. *Information Visualization*, *10*(2), 97-106.

Shipman, F., Hsieh, H., Maloor, P., & Moore, J. (2001). The visual knowledge builder: a second generation spatial hypertext. In *Hypertext and hypermedia* (p. 113-122). New York, NY, USA: ACM.

Starbuck, W., & Milliken, F. (1988). Executives perceptual filters: What they notice and how they make sense. In *The executive effect concepts and methods for studying top managers* (Vol. 35, p. 35-65). JAI Press.

Stelmaszewska, H., Wong, W., Attfield, S., & Chen, R. (2010). Electronic resource discovery systems: from user behaviour to design. In *Nordichi* (p. 483-492). New York, NY, USA: ACM.

Strobelt, H., Oelke, D., Rohrdantz, C., Stoffel, A., Keim, D., & Deussen, O. (2009). Document cards: A top trumps visualization for documents. *Visualization and Computer Graphics, IEEE Transactions on*, *15*(6), 1145-1152.

Weick, K. (1995). *Sensemaking in organisations*. London, England: Sage.

Wise, J., Thomas, J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., et al. (1995). Visualizing the non-visual: spatial analysis and interaction with information from text documents. In *Information visualization* (p. 51-58). Washington, DC, USA: IEEE Computer Society.

Wong, W., Chen, R., Kodagoda, N., Rooney, C., & Xu, K. (2011). Invisque: intuitive information exploration through interactive visualization. In *Chi extended abstracts* (p. 311-316). New York, NY, USA: ACM.

Wright, W., Schroh, D., Proulx, P., Skaburskis, A., & Cort, B. (2006). The sandbox for analysis: concepts and methods. In *Chi* (p. 801-810). New York, NY, USA: ACM.

Yeh, S., & Liu, Y. (2011). Integrated faceted browser and direct search to enhance information retrieval in text-based digital libraries. *International Journal of Human-Computer Interaction*, *27*(4), 364-382.