DIMACS Technical Report 2009-12
November 2009

Supporting Cognitive Models of Sensemaking in Analytics Systems

by

Jason Perry
Dept. of Computer Science
Rutgers University
New Brunswick, New Jersey 08903

Christopher D. Janneck
Dept. of Computer Science
and Engineering
Lehigh University
Bethlehem, Pennsylvania 18015

Chinua Umoja
Dept. of Computer Science
Morehouse College
Atlanta, Georgia 30314

William M. Pottenger
Dept. of Computer Science and DIMACS
Rutgers University
New Brunswick, New Jersey 08903

ABSTRACT

Cognitive science is providing the scientific community with increasingly well-supported models of the mental stages and representations that professional analysts go through in the course of conducting an investigation, be it reactive or proactive in nature. These process models are generally advanced within the field of Sensemaking, because the analyst's primary task can be viewed as "making sense" of a large body of unorganized information. One of the most well-known long-running Sensemaking investigations is that of Pirolli and Card et al. [Pir05] Their resulting model provides an initial basis for our research.

In using these models to improve analytics systems, we have at least two distinct problems: (1) how to use information about the Sensemaking states gained from user interactions with an analytics system to learn the parameters of an effective analysis process, and (2) how to use this knowledge to provide user guidance that results in better human-machine interaction and a more robust investigative process. The answers to these questions lie at the intersection of research in machine learning, knowledge representation, user interfaces and cognitive science, and addressing them requires an end-to-end system perspective.

In this report, we survey these problems and discuss our approach, system design, and experimental design. In particular, we define the Sensemaking model's representation within the software framework, the set of machine learning tasks for learning the parameters of an efficient process, our initial user interface design, the design of the meta-cognitive UI feedback, and finally the design of the initial experiments, including the ground truth which is from an actual solved crime case. We conclude with the insights gained thus far into building interactive systems that support users' cognitive models of Sensemaking.

1 Introduction

Research in Cognitive Science has reached a point of maturity at which research in interactive analytics can reap significant benefits from the insights Cognitive Science has provided into analyst processes. We now have structured theories of analyst cognitive processes that accurately model cost structures, as shown by empirical Cognitive Task Analysis experiments. Research into these models generally falls under the heading of "Sensemaking."

The search is underway for ways to implement these models in analyst software support environments in ways that demonstrably improve the quality of the analyst process. We offer an approach that is realizable with current technology and generalizable to a broad array of cognitive models.

In this report we detail one proposed approach to integrating a Sensemaking model in an Interactive Analytics System, incorporating the cognitive model's representation, machine learning to determine the parameters of efficient analysis processes, and intelligent feedback in the user interface. We will present our approach with Pirolli and Card's model primarily, but it will also be made clear how the method is generalizable to other models. Henceforth we refer to the system we are building as the *Sensemaking Software Support System*, or *S4*.

In Section 2 we give background in the related areas of data analytics and Intelligent User Interfaces. Section 3 describes the Sensemaking model that has been proposed and developed by Russell, Stefik, Pirolli and Card, and further developed by Pirolli and Card et al. In Section 4 we describe the technologies that we believe are foundational to effective support of a Sensemaking model in Interactive Analytics Systems. The subsequent sections discuss the details of the design of our own Interactive Analytics system, named S4, including its machine learning component, user interface, and intelligent assistance mechanisms. We conclude with a description of the initial experimental design and future directions.

2 Background: Data Analytics and Intelligent User Interfaces

This research takes place in the broader context of Data Analytics and Intelligent User Interfaces, and we bring an awareness of developments in these fields to bear on the problem of integrating Sensemaking models into analytics systems.

The process of analysis of data, even foregoing the dissemination and collaboration aspects, presents a difficult task for humans. The challenge and complexity of the analytical tasks at hand are quickly met by the limitations in human cognitive processing. These limitations both involve the difficulties in processing large capacities of information, as well as the natural habits and biases that humans create and engender through the analysis process and daily living [Heu09]. These human limitations compound the difficulty of performing analysis using desired practices and mindsets, and produce the potential for "failures" at the analysis level, even when the necessary facts were collected. Examples of such failures in the realm of military intelligence include the events of the attack on Israel on Yom Kippur, 1973 (which was correctly predicted, but suppressed) and the Falkland Islands, 1982 (which was incorrectly reported for political and monetary expediency) [Clark04].

To date, the majority of research efforts into analytical support have focused on supporting the performance of various analytical tasks. Conversely, limited attention has been given to user interfaces which support the actual analytic process itself [ZM06]. This support may come in a synergistic relationship between the human and computer working collaboratively, with the computer understanding and assisting with their tasks (e.g., reminding them to apply particular techniques), and the human leveraging additional work provided by the computer (e.g., automatic follow-up queries) and being managed through the analytical process [ZM06]. These prominent research thrusts can also be seen as the synergistic approaches through HCI (enhancing human abilities with better interfaces) and Artificial Intelligence (creating more helpful agent assistants) [He07]. The research herein, however, is more closely represented by work on adaptive systems that apply user modeling (e.g., [Santos03], [Santos05]), but with emphasis on process support instead of information exploration.

In short, while intelligence analysts have many techniques to assist in developing thorough and objective products, they, like all of humankind, often fall short of executing an "ideal" analytical method due to a number of cognitive, psychological and cultural factors. Particularly well-studied limitations include cognitive biases such as attention span, failure to generate alternative hypotheses [TK74] and confirmation bias.

It is this context that provides the driving force behind this research: to develop techniques that encourage more efficient human performance of the analytical method. To date our research has focused on the study of, and aim to assist, one particular group whose daily work exemplifies the need for enhanced Sensemaking research and support: the Intelligence Community (IC). "Expected to bring to the table a capacity to draw reasoned and actionable conclusions from scant and conflicting information" "with the cost of failure catastrophically high"[Clark04], the IC is one critical group that will find actionable Sensemaking research immediately useful. Still, applying groundbreaking research within the IC is a daunting task given the number of legal, political, ethical, security and other concerns with information sharing (even assuming complete technical integration). To address these real-world concerns, our Sensemaking research is rooted within a higher-level framework called DI HOPE KD [Jewett08] [Cru08], which is exploring the following aspects of Interactive Automation [NSF07]:

- Distributed - focusing on algorithms that operate in hybrid (non-horizontal, non-vertical) distributed data environments [NLP09] [Li07]

- Interactive - leveraging synergistic efforts between humans and machines (discussed herein)

- Higher-Order - investigating the utility and application of links that leverage higher-order paths in graphs [GLP09] [GPY06]

- Privacy-Enhancing - sharing relevant linking meta-information while preserving the privacy of specific data [NLP09]
- Knowledge Discovery - revealing hidden and discovering new knowledge [GPY06]

Our Sensemaking work will be incorporated into DI HOPE KD with the goal of facilitating widespread utilization of these merits and capabilities. The promise of the Sensemaking model is to provide an explicit representation of the unseen states of the analysis process, providing the means to identify the precise points at which these cognitive limitations hinder the investigative process, so that targeted feedback can be given in overcoming them.

3 The Sensemaking Model

All Sensemaking models attempt to give a structured account of the nature of critical thinking, which is key to any successful investigation. In the context of this work, we refer to the definition of critical thinking provided by David Moore, intelligence officer and academic for the National Security Agency, rephrased from Paul and Elder (2004):

"Critical thinking is a deliberate meta-cognitive (thinking about thinking) and cognitive (thinking) act whereby a person reflects on the quality of the reasoning process simultaneously while reasoning to a conclusion. The thinker has two equally important goals: improving the way she or he reasons and coming to a correct conclusion." [Moore07]

The structures of a Sensemaking model serve to shed light on the types of meta-cognitive acts that are seen as necessary parts of the Sensemaking process. Sensemaking models typically consist of a number of stages and a pattern in which the analyst enters and exits these stages over the course of an investigation. Each stage suggests modes of critical thinking that are likely to be active at that point in the investigation, as well as implying the type of data-centered tasks that an analyst would be likely to perform in that stage.

One of the most well-known models of Sensemaking is that of Pirolli and Card [Rus93]. A reproduction of the diagram Pirolli and Card's Sensemaking model is shown in Figure 1.

Through Cognitive Task Analysis, Pirolli, Card, and their collaborators have found that as analysts gain expertise they learn to create structured representations of knowledge. These representations are the "artifacts" represented in the square boxes in Figure 1. As the investigation progresses, the analyst can be seen as developing and revisiting a more and more structured set of artifacts, until finally he or she arrives at a "presentation": a concise representation of the answer that was sought that also convincingly justifies that answer.

The nested-loop structure of this model captures the iterative and cyclic path that analysts take amongst their data sources and artifacts in the course of real investigation. For any given analytical task, an analyst would be seen to blaze a structured trail through these stages, visiting them in a different order and remaining in each stage for varying amounts of time. Furthermore, Pirolli and Card's model differentiates between "top-down" and "bottom-up" investigative tasks. In brief, a "top-down" task is one in which the analyst is moving from theory to data, and a "bottom-up" task is one in which the analyst is

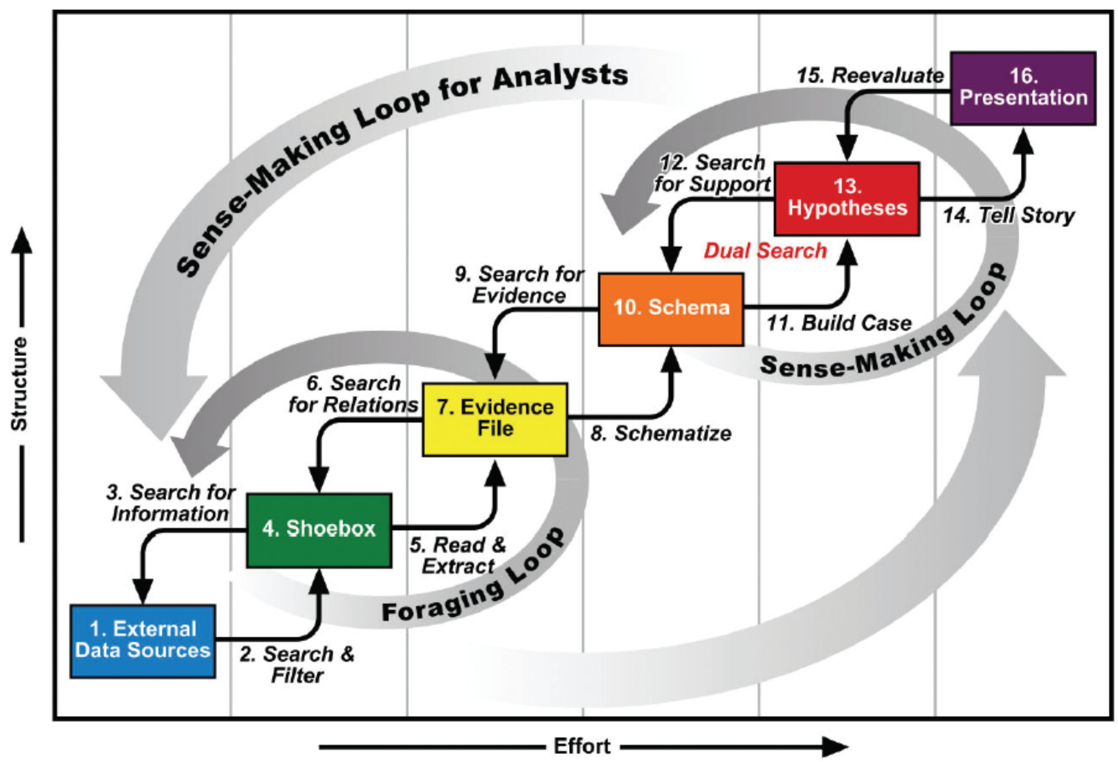

Figure 1: The notional model of Sensemaking.

moving from data to theory. The process of moving "bottom-up" through the Sensemaking stages can be seen as constructing a more and more structured representation of case knowledge, while it often becomes necessary to go "top-down" and revisit lower-level structures to supply missing information. We believe the bottom-up/top-down distinction is key in the analysis in cognitive states and processes, and this issue will be seen to drive multiple design decisions of our support system.

Cognitive task analysis research has identified leverage points within the Sensemaking model at which the cognitive limits of analysts are reached and which represent critical points in the exploration/exploitation tradeoff curve [Pir05]. This is strong evidence for the feasibility of experimentally determining model parameters that represent efficient paths through the model's stages. In other words, if expert analysts are those who have learned through experience to make tradeoffs that maximize human cognitive capabilities, then capturing similarities in the structure of these experts' Sensemaking processes will define a generalized process model which represents a highly effective human Sensemaking process.

4 Key Components of Sensemaking Support

At the core of our approach is the idea that an interactive analysis system can use data from user interactions to infer a high-level knowledge of the analyst's progress within a Sensemaking process, and then use this high-level knowledge to provide feedback that encourages a highly effective route through the process. In other words, the software should know enough about what the user is doing to be able to support a Sensemaking profile that provides high efficiency and minimizes errors due to known human cognitive limits.

One can see how this higher-level knowledge might be obtained through machine learning. One such "brute-force" machine learning approach might be to record the sequence of user interactions in an analysis system; have a Sensemaking expert, in consultation with the analyst, manually label the transitions between Sensemaking stages that were passed through; and train with these to learn a sequence model, such as an HMM (Hidden Markov Model) or Markov Net [Rab89], that can predict when a transition between Sensemaking states most likely occurs. Experience in machine learning shows, however, that this would produce a large number of parameters which in turn would require significant training data - something difficult to obtain in this domain. With insufficient training data the resulting model would be too sensitive to variations in users' investigative styles, and as a result, generalize poorly to new users. Such a setup would be unlikely to provide any real knowledge about the Sensemaking process. Therefore, in our system, we have taken an approach that provides us with more detailed knowledge of the analyst's process from the user interface itself, and then machine learning is used to model the progress of an expert analyst within this particular framework.

In order to gain high-level knowledge of the analyst's Sensemaking process, we need to determine a mapping from the level of raw user interactions up to the higher-level abstractions of the Sensemaking model. In representing Pirolli and Card's Sensemaking model there are three primary classes of objects: Stages, Artifacts, and Data Tasks. The logical dependency

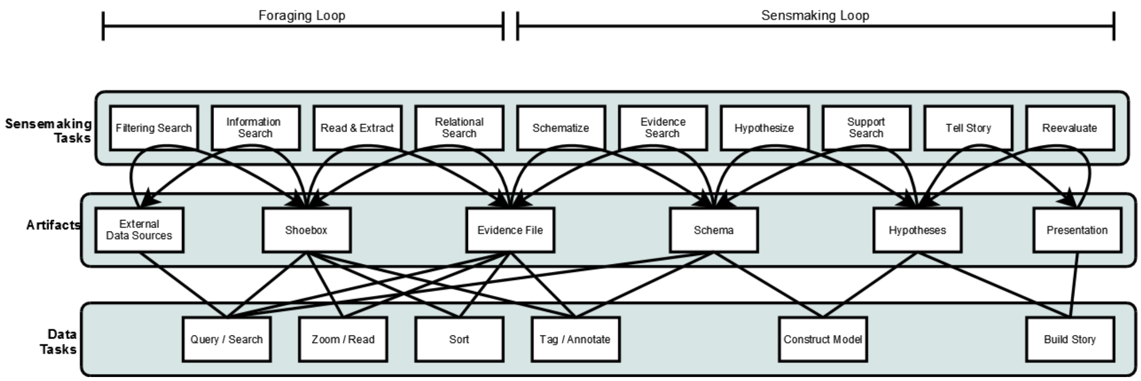

Figure 2: The mapping of task types to Sensemaking stages.

structure of these classes can be encoded in structure of links, producing a set of hierarchical relationships. Pirolli and Card's model does not strictly specify which specific data tasks can be performed in which stages, giving us some freedom in tailoring this mapping to the specific functions our system provides.

Our mappings, in terms of broad categories of data tasks, are depicted in Figure 2, which takes the form of an abstraction-decomposition matrix. The Sensemaking tasks are at the highest layer of abstraction, the artifacts at the next lower layer, and the data tasks at the lowest layer. Our data tasks were chosen from those collected through anthropomorphic studies of analyst processes [Rob08] [GZ08] [ITC08]. We have grouped the data tasks into six major categories: Query, Zoom, Sort, Annotate, Schematize, and Build Story. The actual set of data tasks provided for in our user interface is much more fine-grained; the detailed mappings of data tasks to artifacts are expressed in the design of the user interface and the actions it permits at a given stage. The execution of these data tasks can be recorded directly from the user interface as a linear history, and this history can be analyzed to provide more insight into the analyst's progress through the process, as we will discuss in the section on Machine Learning.

5 The Design of the Sensemaking Software Support System

In developing our system, S4, as an integrated analytics system implementing the aforementioned components of Sensemaking support, we have focused our design around the following key drivers:

- Should leverage existing analytics technologies at the data storage/mining/query level, if possible be based on an existing platform in use in research or real-world analytics
- Sufficient control over the features of the analytics environment for well-controlled experiments with broadly applicable results.

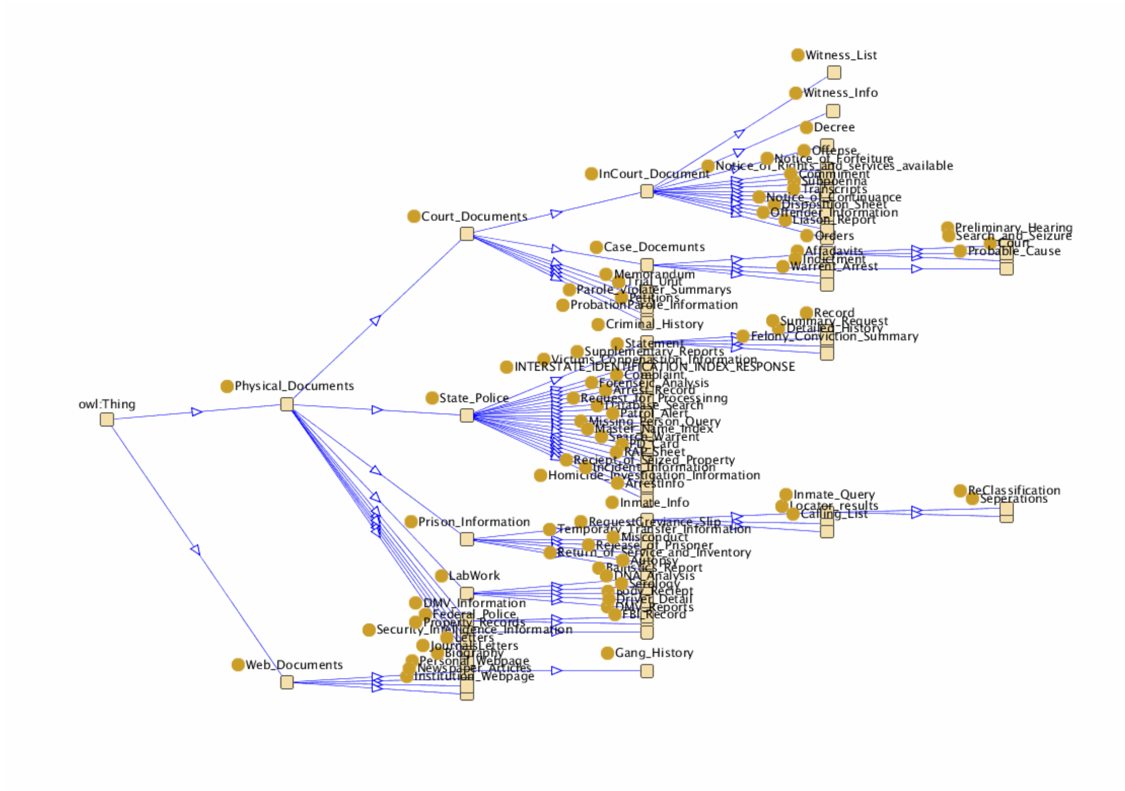

Figure 3: A visual representation of the ground truth OWL ontology.

- Sufficient freedom for the analyst to be able to make full use of his or her existing expertise and to follow as 'natural' an investigative process as possible.
- Modularity, allowing for the Sensemaking support itself to be adapted to different models as well as analytics systems
- Inclusion of a real-world ground truth data set for experimental purposes
- Short implementation time

S4 will initially support investigative analysis over knowledge bases that are stored in Blackbook. Blackbook, developed by IARPA under the KDD program, is a J2EE-based Web Service provider being developed to support investigative analysis by leveraging Semantic Web technologies. Blackbook supports a large number of query types that are federated across multiple data sources. Blackbook stores its knowledge base using the Resource Description Format (RDF), a W3C standard that represents knowledge facts as triples, and can be readily translated into a graph, XML, or other formats. Using Blackbook readily provides support for our desired design factors of modularity and rapid development on an existing analytics platform.

As the foundation of a ground truth for these experiments, we have access to a data set provided by the Bethlehem, PA police department. It is the complete record of a complex solved case, including hundreds of documents, and specifying a well-defined series of relationships that need to be extracted to solve the case. Still, in order for S4 to make use of these relationships, the entities and relationships found in the data must be semantically encoded in an appropriate format, transforming it into a "ground truth" data set. The target format for this ground truth, which originally consisted of a collection of PDF documents and their OCR outputs, is the Blackbook RDF format. The development of this initial ground truth has required several knowledge engineering steps. First, we studied the various document types in the corpus and crafted an OWL ontology by which to organize the data. A graph representation of a subset of the ontology is shown in Figure 3. Then we evaluated the generated OCR from the source documents, cleaning the data where needed and verifying the integrity of the generated text. After refining the ontology again (now defining a "class" for each document type), we then developed scripts to extract the textual data from the generated text into RDF files representing each document class. These files are then loaded into Blackbook and again verified for relational integrity through the Blackbook access methods.

6 The Artifact-Based Sensemaking User Interface

Our initial design for a User Interface to support Sensemaking provides a very explicit representation of key features of the Sensemaking process, wherein there is a user interface widget corresponding to each artifact in the model. The basic mode of operation is for the analyst to work within a single artifact widget, with the ability to switch to either the previous or following widget (represented as a window in the user interface) as specified by the linear order of the Sensemaking model. Each widget provides the ability to enhance and refine the primary data structures that result from the application of its respective artifact, and supports propagation of data structures from the current artifact widget to adjacent ones.

Several factors currently motivate this interface design, which is in effect the direct adaptation of a Sensemaking model as the driving interface metaphor, which is novel in current literature. Regarding theory, we posit that if Pirolli and Card's Sensemaking model is an accurate picture of common cognitive states and knowledge structures of analysts, then an interface that explicitly represents these knowledge structures will produce "naturally" efficient analysis procedures. In addition, the visual representation of artifacts should serve to stimulate novice analysts in developing the same style of structured mental representations that have been observed in expert analysts. Pragmatically, an interface that directly represents the primary data artifacts encapsulated by the actions invokable upon them provides a strong suggestion to the artificial intelligence components as to the current state of the user's mental process per this Sensemaking model. For example, if a user is using the Query widget extensively, this would suggest a current mental goal of information foraging and operating in the lower-left corner of the Foraging Loop (per Figure 1). In addition, restricting the user's widget-transfer path options makes the artificial intelligence challenges discussed in

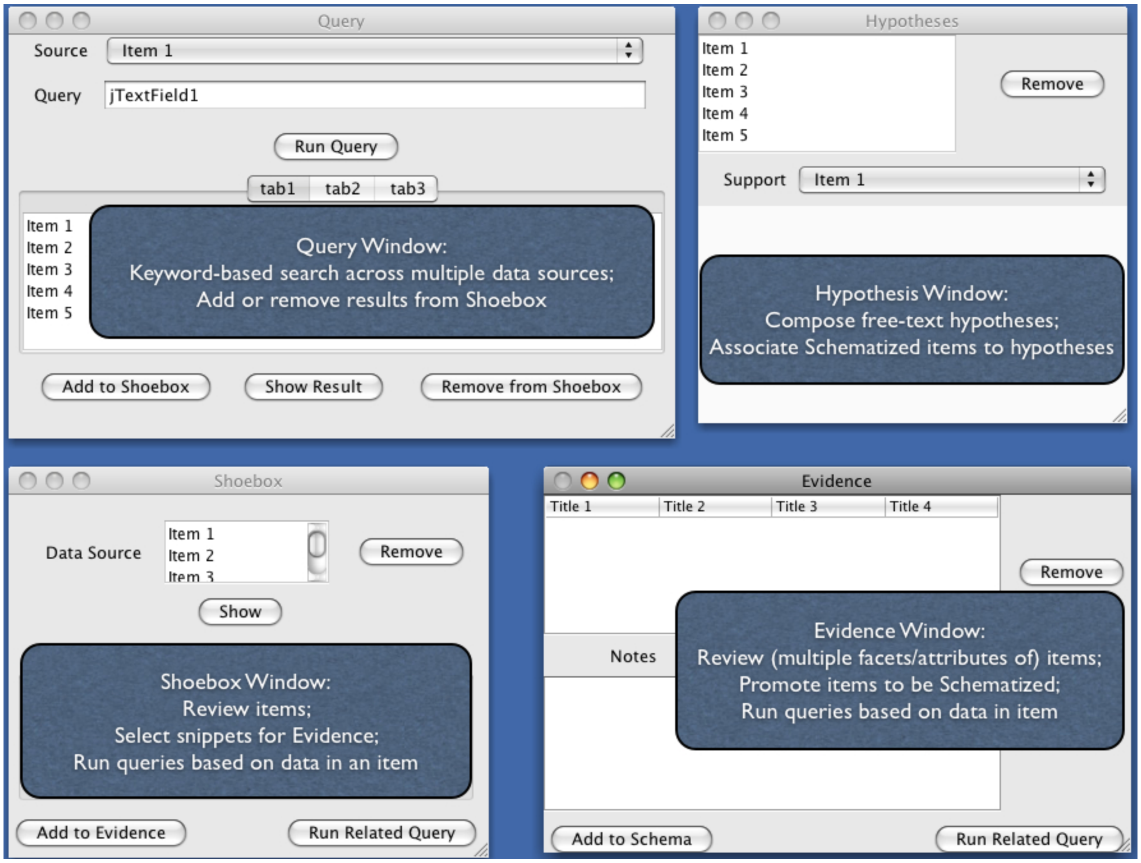

Figure 4: Four artifact widgets from the Sensemaking User Interface.

Section 7 significantly more tractable.

In accordance with this motivation, we are implementing a user interface that explicitly represents the artifacts of the Sensemaking model, in the form of one interactive widget for each artifact. Figure 4 shows prototypes of four of the artifact widgets: Query, Shoebox, Evidence File, and Hypotheses. The following is a description of each of the artifact widgets in the user interface.

Query - the primary interface between the user and their available data sources (Web, Database, File system and other), the Query artifact widget resembles a traditional keyword-based search application. Upon selecting a data source and entering the search phrase, a list of results are displayed which can then be viewed and, if deemed potentially relevant to the analysis at hand, added to the Shoebox.

Shoebox - the initial cache of collected source documents, the Shoebox allows the user to review the contents of each in greater detail, and move important pieces forward into the Evidence File. Items no longer seen as important can also be removed from the Shoebox.

Selecting a Shoebox item from the list displays the contents of that item, and also allows the user to add the entire item, or only a snippet, into the Evidence artifact widget. Users can also use text from Shoebox items to drive additional queries on data sources (in the Query artifact widget), in a query-by-example-like approach to return similar or related results.

Evidence File - allows the user to review the snippets and documents contained herein and propagate related items to the Schema artifact widget. Similar to the Shoebox artifact widget (as a second-tier information item filter), the user can add selected items to the Schema artifact widget, remove them from the Evidence artifact widget, or run a query based on the contents of the item upon the Shoebox artifact widget(searching for other relevant pieces of data).

Schema - facilitates the application of multiple visual analysis techniques on accumulated Evidence items to reveal relationships amidst information that can be used to support one or more hypotheses. The user selects various items to review and a visualization within which to view them, and can capture these resulting visualizations and relationships into schematized items, which can then be added to the Hypothesis artifact widget.

Hypothesis - allows the user to compose free-text hypothetical statements, and associate schematized items with them. A hypothesis can have one or many schematized items associated with it.

Presentation - allows the user to organize one or more hypotheses along with supporting data, in order to tell a story that justifies the chosen hypotheses. These representations can then be organized into a linear series of bullet-point slides which can then be exported for presentation to others.

By directly implementing a Sensemaking model as tangible widgets in the S4 interface, we seek to investigate the effect of supporting the overarching analytical process. This approach is distinct from current research efforts, whose primary focus is to enhance one or more specific tasks performed during analysis (e.g., enhanced queries and filters [Stumpf08], evidence marshaling [SV08] [Wright06], intelligent data visualization [Wen07] [Wen08].) These approaches are not exclusive, however, as follow-on work will begin to integrate and evaluate our efforts with process-centric support with other task-centric support. This interface model also accommodates a modular design including automated process suggestions (see Section 7) and interface techniques applied to these artifact widgets to guide a user's process (see Section 8).

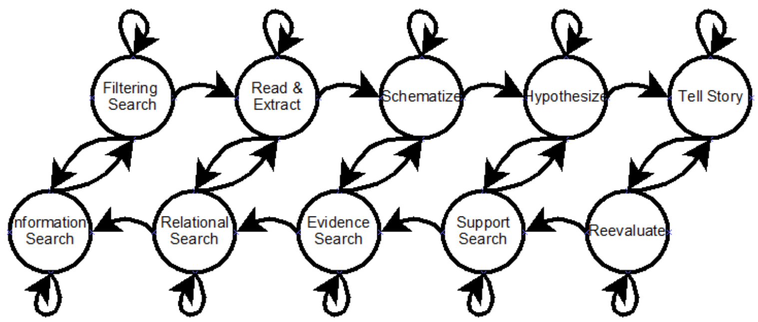

Figure 5: A diagram of an HMM for Sensemaking States.

7 Machine Learning for Human Process Optimization

The goal of machine learning within S4 is to use the available information from the user interface to determine the parameters that characterize an efficient Sensemaking process, given the parameters of a given data set. In particular, we want to use the record of interactions of expert analysts on real case data sets to learn a model of transitions between the artifact widgets that represents an effective Sensemaking process. This will enable the system to guide future users toward a more efficient approach to Sensemaking, as well as providing insight into which features of Pirolli and Card's model are truly critical for characterizing an effective analysis procedure.

One well-known machine learning model for modeling time-dependent processes with multiple states is the Hidden Markov Model, (HMM) [Rab89]. In our system, the states of the HMM correspond directly to the artifact states of the Sensemaking model, and the observation sequence corresponds to the sequence of data tasks as recorded from interactions with the User Interface. Here the state transitions are known, since they are explicitly represented in the User Interface; so the HMM is not used to compute the most likely sequence of state transitions from an observation sequence. Instead, we use the HMM to compute the likelihood of a transition to a different artifact widget at any given point in the process.

In our system, the HMM will be trained on expert analysts who are known to have a very efficient analytical process. The training procedure is the same as in any other application of an HMM and proceeds as follows: after an initial training period with the system, the expert analysts in the experimental group will use the system without assistance to solve multiple ground truth cases, as specified in the experimental design section. Each record of state transitions and data tasks for a single solved case will make up one training sequence for the HMM model of the Sensemaking process. From the sequence of state transitions and data actions, the parameters of the HMM's transition and observation sequence probability distributions will be learned by means of the well-known Baum-Welch algorithm [Rab89]. These probability distributions, being generalized over multiple investigators and data sets, will represent patterns of analytic activity that make highly effective use of the available data and knowledge representation structures.

Of key importance in the success of this procedure is the selection of features recorded in the observation sequence, as these become the features of the training set. As mentioned above, the primary feature set will simply be an encoding of each of the types of data tasks that can be recorded in the user interface. However, we also plan to use explicitly constructed aggregate features that are likely to have a strong causal link to state transitions. We anticipate this will produce significantly improved performance over using the raw input stream alone. Potential aggregate features are the number of data items stored in the current window and the count of data tasks of each type.

It is also possible that a classical HMM may not be sophisticated enough to capture the essence of a wide range of effective paths through the Sensemaking process. One reason for this is that the Markov chain has no explicit record of how long it has remained in its current state. Though the self-transition probabilities of HMM states implicitly model an exponential density for the duration of remaining in one state, this may not be the most appropriate approach for the behavior we wish to model. In principle, at least, an HMM can remain in one state for an arbitrary number of time steps. Initial experiments will be used to determine whether it is necessary to use more sophisticated models such as non-stationary Hidden Markov Models [SK95].

Once the model parameters have been learned, the HMM inference algorithms can be used to monitor and guide the Sensemaking process. While an analyst is carrying out an investigation using the system, the system will again record the sequence of user data tasks. After each new action, the HMM is used to compute the probabilities of three events: the probability of the user remaining in the same artifact state for the next action, the probability of switching to the adjacent higher-level artifact widget, and the probability of switching to the adjacent lower-level artifact widget. These probabilities can be computed by a direct product, as the state transitions (and model parameters) are known. Furthermore, this an efficient process as the probability of the whole sequence does not need to be recomputed each time.

These probabilities can tell us several things. If the probability of a transition (in either direction) becomes much higher than the probability of remaining in the current state, it may indicate that the analyst has remained in one state past the point of diminishing returns and could use some assistance in either moving on to a higher level of organization (i.e., the next higher-level artifact) or gathering more supporting evidence (i.e., the next lower-level artifact). For example, the analyst may be continuing to collect new search results past the point where it will be practicable to follow up on so many results. Another possible scenario is where the analyst has gathered a set of related evidence snippets, and the number of those snippets has surpassed the limit of short-term memory. At this point it may be advisable to group the evidence snippets in some schema, to produce a higher-level organizational structure that serves as a memory aid.

In addition, if the user makes a transition between artifact widgets at a point when the HMM indicated that the probability of such a transition was very low, assistance can be given to the effect that the user should spend more time exploiting the information available in the artifact widget he or she has just left. Actual thresholds for the probability values will be determined experimentally and can be different for each artifact widget. The nature of the specific types of meta-cognitive assistance to be provided are described in the next section.

8 Meta-Cognitive Assistance for Human Process Optimization

The Sensemaking support system will implement meta-cognitive support through interface techniques that encourage the user to follow the efficient process model of analysis as per the system's learned model. The most common interface techniques to support the user through complex tasks to date typically involve activities better suited to follow a linear series of steps [Dryer97], inferring user needs and simplifying (i.e., removing controls from) the interface [MBB02], or allowing end-users to create "guides" of repetitive procedures [Spot07]. While these techniques do provide some suggestions for guiding users towards particular simple actions (e.g., video tutorials enhanced with text, highlighting and deictic graphical overlays), these approaches do not generally translate well to the open-ended non-linear analytic process, and do not address encouraging adherence to process models other than by repetition.

We seek both to leverage interface techniques from other domains and also to develop new ones in the context of this work. [CKB08] provides a survey of four classes of techniques useful in the visualization and manipulation of amounts of information that exceed the available screen space. Overview+Detail techniques provide a spatial separation between a high-level aspect for context and low-level view for intricate operations. This technique, stemming back to the video game Defender in the 1980's, is now widely used in applications, including many Web-based mapping sites. Zooming interfaces allow the user to traverse along the z-plane, moving between overview and detail-type views across time. Focus+Context interfaces distort the information space to give greater space to the item or area of current interest (i.e., under "focus"). Fish-eye views (notably adopted in the Mac OS X Dock), hyperbolic trees, and code folding in text editors are all examples of this technique. Cue-based techniques do not modify the size of objects but rather their rendering properties to distinguish between the objects in focus and their surrounding context. Blurring contextual items and using decorators to hint at objects out of view are examples of these techniques.

We consider these techniques not only for their ability to support certain kinds of task actions, but also for their potential to discourage others. For example, a Fish-eye mechanism could be adopted to both encourage cognitive focus on the artifact widget at hand (i.e., in focus), while also being used to discourage continual action in the same artifact widget (e.g., by moving the fish-eye-focus towards another artifact widget).

We propose the following interface techniques as a sample of the kinds of techniques to be investigated in the present system, to support user guidance:

- Sensemaking Overview Process widget: displays the Pirolli and Card model (per Figure 1) including distributions of time spent, highlighting phases needing more time

- Zoomable artifact widgets: ”zoom into” (make larger) a specific artifact widget to encourage use; zoom-out (make smaller) to discourage use
- Suggestion “Box”: reserved space for guidance messages from system, and a button to bring a suggested activity into focus
- Suggestion “Bubble”: guidance messages relevant to the artifact widget they refer to appear in a “pop-up” fashion
- Highlighting/bouncing/pointing to artifact widgets: cues to encourage usage of a particular widget
- Abstracting artifact widgets: hiding details of artifact widgets to discourage use

9 Experimental Design

The empirical evaluation of the Sensemaking Support System will be achieved through an informal pilot study, a formal training period and a formal experimental session.

We are planning on holding three human-participants activities in the near term. All will follow a similar format, including quantitative (e.g., task completion time, Sensemaking phase durations, number of queries performed, percentage of Shoebox/Evidence/Schematized items propagating to the result) and qualitative metrics (e.g., Likert scale ratings assigned to personal and system evaluation questionnaires, the NASA TLX for workload evaluation).

The first pilot study will be focused on collecting data and system parameter adjustment. All participants will be regional police detectives and officers. Participants will be asked to generate an analysis report on the primary actors and activities from an unclassified data set. One potential data set in development describes entities and activities surrounding the terrorist attacks on Sept. 11, 2001. Data (both that captured by the system and from qualitative surveys) will be used as training data for the machine learning model as well as to suggest which feedback techniques and configurations are most promising.

The second activity will be a training session for the experimental participants. All subjects will be regional police detectives. The setup, source data, activities and data collection will be similar to the pilot test, with the following exceptions: 1) all participants will undergo a formal training process; 2) only the treatment group will use a set of pre-selected guidance techniques and parameters while undertaking training tasks; 3) only the treatment group will be allowed to enable, disable and reconfigure the guidance techniques at the end of the training period. Activity logs and qualitative surveys will again be captured, and will be used to fine-tune the techniques and parameters used for the final experiment.

The third activity will formally compare the baseline group (using the artifact-based user interface for analysis) with a treatment group (using the same interface, with the guidance techniques and analytical assistance engine enabled). The data set will be a ground truth consisting of police documents pertaining to a closed murder case from the Bethlehem Police Department. This set is a complete record of a complex solved case, including hundreds of documents and specifying a well-defined series of relationships that need to be revealed to solve the case. Participants will be instructed that a crime has occurred, and their task is to generate an analytical report that contains 1) the nature of the crime, 2) the primary actors and their relationships, and 3) the suspect. To further simulate a real-life investigation, all users will employ a modified query system that requires analysts to request specific documents, as if they were about to collect the information themselves. For example, the user will not be able to search for all documents containing the terms "Joe Smith," but the user will be able to search for an interview with Joe Smith. We hypothesize that the treatment group will perceive a better analytical result when compared to the baseline, and that a specific range for the parameters that regulate the feedback will have a more significant perceived effect on the interface techniques employed by the analyst.

Due to the logistical difficulties in co-locating regional law enforcement detectives for training and experimental purposes, we will travel to participating detectives' offices and facilitate the activities with smaller groups. Participants will all be able to connect to the experimental system, planned to be deployed at the Bethlehem Police Department, securely across the network. The data from the small groups will be combined before initiating analysis.

10 Conclusion

The near-term work is to carry out the experiments as described above. The initial user interface and architecture described above can also be seen as a source of data for the long-term work.

One apparent long-term goal is to build an analytics framework that can model and guide the analyst's Sensemaking process even in user interfaces that do not explicitly model the structures of the Sensemaking model. Such a system would be able to give genuinely "enlightened feedback" to help expert analysts refine and optimize their investigative process on whatever analytics platform they may use. The theory and know-how to be gained from the current system is a necessary component in the realization of such a system. The successful integration of Sensemaking models into analytics systems will represent a new type of computational steering for data analytics, where the actual cognitive states of the human user are used to steer an extended computation taking place across multiple distributed data sources.

The design presented here is one instance of a more general methodology wherein a cognitive workflow model is embodied in the machine by means of probabilistic graphical models, and then learning is done on those models using user interaction data in order to determine efficient workflows. We believe this is an approach that can be profitably utilized in the design and testing of many workflow models as well as the implementation of more advanced analytics system interfaces. For example, we believe that the alternative frame-based Sensemaking model by [Klein06] can be modeled in this way.

11 Acknowledgments

The authors wish to thank Rutgers University, the National Science Foundation and the Department of Homeland Security. This material is based upon work partially supported by the National Science Foundation under Grant Numbers 0703698 and 0712139 as well as by the Department of Homeland Security Science & Technology Directorate. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the Department of Homeland Security or Rutgers University.

We are also grateful for the help of other co-workers, family members and friends. Co-authors W. M. Pottenger, C. D. Janneck, and J. Perry also gratefully acknowledge the continuing help of their Lord and Savior, Yeshua the Messiah (Jesus the Christ) in their life and work.

References

[CKB08] Cockburn, A., Karlson, A. and Bederson, B.B. A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. ACM Computing Surveys, 41(1), 2008, article 2.

[Clark04] Clark, R.M. (2004) Intelligence Analysis: A Target-Centric Approach. CQ Press.

[Cru08] Cruikshank, D. (2008) Offering New HOPE in the Balance Of Security and Civil Liberties. Online at `http://nsf.gov/discoveries/disc_summ.jsp?cntn_id=112771&org=IIS&preview=false`

[Dryer97] Dryer, D.C. Wizards, guides, and beyond: rational and empirical methods for selecting optimal intelligent user interface agents. In Proc. International Conference on Intelligent User Interfaces, 1997 pp. 265-268.

[GLP09] Ganiz, M. C., Lytkin, N. I. and Pottenger, W. M. (2009) Leveraging Higher Order Dependencies Between Features for Text Classification. In the Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD). Bled, Slovenia, September.

[GPY06] Ganiz, M. C., Pottenger, W. M. and Yang, X. (2006) Link Analysis of Higher-Order Paths in Supervised Learning Datasets. In the Proceedings of the Workshop on Link Analysis, Counterterrorism and Security, 2006 SIAM Conference on Data Mining. Bethesda, MD, April.

[GZ08] Gotz, D. and Zhou, M.X. (2008) Characterizing Users Visual Analytic Activity for Insight Provenance. In Proc. IEEE Symposium on Visual Analytics Science and Technology, pp. 123-130.

[He07] He, D. Brusilovisky, P., Grady, J., Li, Q. and Ahn, J-W. (2007) How Up-to-date should it be? the Value of Instant Profiling and Adaptation in Information Filtering. In Proc. IEEE/WIC/ACM Conf. on Web Intelligence 2007, pp. 699-705

[Heu09] Heuer, R.J., Jr. (1999) Psychology of Intelligence Analysis. Center for the Study of Intelligence, Central Intelligence Agency (CIA), Washington, DC.

[ITC08] Isenberg, P., Tang, A., and Carpendale, S. (2008) An Exploratory Study of Visual Information Analysis. In Proceedings of CHI 2008, pp. 1217-1226.

[Jewett08] Jewett, Barbara (2008) Mimicking Human Intuition. In NCSA Access, Fall 2008, pp. 20-23. Also available online at `http://www.ncsa.illinois.edu/News/Stories/Intuition/`.

[Klein06] Klein, G., B. Moon and R.R. Hoffman, "Making Sense of Sensemaking 1: Alternative Perspectives," *Intelligent Systems*, vol. 21, no. 4, 2006, pp. 70-73.

[Li07] Li, S., Janneck, C., Belapurkar, A., Ganiz, M. C., Yang, X., Dilsizian, M. D., Wu, T., Bright, J. and Pottenger, W. M. (2007) Mining Higher-Order Association Rules from Distributed Named Entity Databases. In the Proceedings of the IEEE International Conference on Intelligence and Security Informatics 2007 (ISI 2007), New Brunswick, NJ, May.

[MBB02] McGrenere, J., Baecker, R.M. and Booth, K.S. An evaluation of a multiple interface design solution for bloated software. In Proc. SIGCHI conference on Human factors in computing systems, 2002, pp. 164-170.

[Moore07] Moore, D.T. (2007) "Critical Thinking and Intelligence Analysis". Occasional Paper No. 14, NDIC Press.

[NLP09] Nikolov, A., Li, S. and Pottenger, W. M. (2009) Privacy-Enhancing Distributed Higher-Order ARM. In the Proceedings of the Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining, Sparks, NV, April.

[NSF07] Pottenger, W.M. (2007) Visual Analytics for Steering Large-Scale Distributed Data Mining Applications. NSF Award No. 0712139. Abstract online at `http://nsf.gov/awardsearch/showAward.do?AwardNumber=0712139`.

[Pir05] Pirolli, P. and Card, S. "The Sensemaking Process and Leverage Points for Analyst Technology as Identified through cognitive Task Analysis." In *Proceedings of 2005 International Conference on Intelligence Analysis.*

[Rab89] Rabiner, L. R., "A Tutorial on hidden Markov models and selected applications in speech recognition", *Proc. IEEE,* Vol. 77, No. 22, February 1989, pp. 57-285.

[Rob08] Robinson, A.R. (2008) Design for Synthesis in Geovisualization. PhD Dissertation, Pennylvania State University.

[Rus93] Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K., "The Cost Structure of Sensemaking". Paper presented at *the INTERCHI '93 Conference on Human Factors in Computing Systems*, Amsterdam, 1993.

[Santos03] Santos Jr., E., Nguyen, H., Zhao, Q. and Wang, H. (2003) User modeling for intent prediction in information analysis. In Proc. 47th Annual Meeting for the Human Factors and Ergonomics Society (HFES-03), pp. 1034-1038.

[Santos05] Santos Jr., E., Zhao, Q., Nguyen, H. and Wang, H. (2005) Impacts of User Modeling on Personalization and Information Retrieval: An Evaluation with Human Intelligence Analysts. In Evaluation of Adaptive Systems, workshop of 10th International Conference on User Modeling (UM2005), pp. 27-36.

[Scholtz08] [Scholtz] (2008) Progress and Challenges in Evaluating Tools for Sensemaking. In Proceedings of CHI 2008 Sensemaking Workshop.

[SV08] Shrinivasan, Y.B. and van Wijk, J.J. (2008) Supporting the analytical reasoning process in information visualization. In Proc. ACM CHI 2008, p1237-1246.

[SK95] Kim, B., and Kim, J. H., "Nonstationary Hidden Markov Model". In *Signal Processing, 46,*, pp. 31-46, 1995.

[Spot07] Spotfire DecisionSite. (2007) Software. Website: http://www.spotfire.com

[Stumpf08] Stumpf, S., Sullivan, E., Fitzhenry, E., Oberst, I., Wong, W-K and Burnett, M. (2008) Integrating Rich User Feedback into Intelligent User Interfaces. In Proc. ACM Conf. on Intelligent User Interfaces (IUI) 2008, p50-59.

[TK74] Tversky, A., and Kehneman, D., Judgment under uncertainty: Heuristics and Biases *Science, 185,* 1124-1131, 1974.

[Wen07] Wen, Z., Zhou, M. and Aggarwal, V. (2007) Context-aware, adaptive information retrieval for investigative tasks. In Proc. ACM Conf. on Intelligent User Interfaces 2007, p121-131.

[Wen08] Wen, Z. and Zhou, M.X. (2008) An optimization-based approach to dynamic data transformation for smart visualization. In Proc. ACM Conf. on Intelligent User Interfaces 2008, p70-79.

[Wright06] Wright, W., Schroh, D., Proulx, P., Skaburskis, A. and Cort, B. (2006) The Sandbox for analysis: concepts and methods. In Proc. ACM CHI 2006, p801-810.

[ZM06] Zhou, M.X. and Maybury, M. Intelligent user interfaces for intelligence analysis. In Proc. ACM Conf. on Intelligent User Interfaces 2006, p. 16.