CRICTO: Supporting Sensemaking through Crowdsourced Information Schematization

Haeyong Chung* Sai Prashanth Dasari* Santhosh Nandhakumar* Christopher Andrews†

*University of Alabama in Huntsville †Middlebury College

ABSTRACT

We present CRICTO, a new crowdsourcing visual analytics environment for making sense of and analyzing text data, whereby multiple crowdworkers are able to parallelize the simple information schematization tasks of relating and connecting entities across documents. The diverse links from these schematization tasks are then automatically combined and the system visualizes them based on the semantic types of the linkages. CRICTO also includes several tools that allow analysts to interactively explore and refine crowdworkers' results to better support their own sensemaking processes. We evaluated CRICTO's techniques and analysis workflow with deployments of CRICTO using Amazon Mechanical Turk and a user study that assess the effect of crowdsourced schematization in sensemaking tasks. The results of our evaluation show that CRICTO's crowdsourcing approaches and workflow help analysts explore diverse aspects of datasets, and uncover more accurate hidden stories embedded in the text datasets.

Keywords: Visual text analytics, sensemaking, crowdsourcing.

1 INTRODUCTION

Various computational approaches have been developed to facilitate the extraction of useful structured information from the veritable mountains of text data being collected. For example, several NLP algorithms, such as topic modelling [1], named entity recognition [2], co-reference resolution [3], relation extraction [4], sentiment analysis [5], etc., have been used to automatically discover and relate semantic information from an unstructured collection of documents.

However, sensemaking of text documents utilizing these existing computational approaches remains challenging, and it is still difficult to form effective semantic meanings and connections from multiple documents depending solely on the algorithms [6, 7]. One of the main reasons is that sensemaking of a large volume of text-based datasets is fundamentally a "cognitively-intensive" task [8, 9, 10, 11] and the human analyst is required to conceptualize a growing body of data, often through the use of judgment and intuition to identify important information and draw conclusions.

Crowdsourcing platforms such as Amazon's Mechanical Turk [12] can be a potentially promising solution for these challenges, since they enable multiple human workers to accomplish cognitively demanding tasks that are complex or challenging for a single analyst (or even several). Thus, the marriage of crowdsourcing with visual analytics can address the growing problem of analyzing data by dividing the analytics process into smaller *microtasks*, which can then be completed by a large number of *crowdworkers*. Previous visualization and data exploration systems have made use of crowdworkers to elicit diverse and useful explanations and annotations in support of the analytic process [13, 14, 15, 16]. These systems could leverage innate human abilities to illuminate patterns, trends and outliers in datasets. Findings from these systems have shown promise that crowdsourcing can be used to gain collaborative insights from datasets and visualizations.

However, it is difficult for existing crowdsourcing data analysis approaches to be applied directly to the analysis of textual data, which requires the somewhat complex processes of foraging for evidence and synthesizing it into explanations and summaries. These time-consuming tasks will likely challenge crowdworkers who (a) have little knowledge of, or experience with, text analysis; or (b) can only commit to analysis work for very short periods of time e.g., 30 minutes. Additionally, although human workers are indeed capable of generating high-quality, insightful microtask results, it remains difficult to combine and synthesize the diverse analysis results from a larger number of crowdworkers into cohesive insights. These problems are certain to be exacerbated in the case of larger document datasets and larger pools of workers.

This paper describes CRICTO (CRowdsourcing Information sChematization TOol) (Fig. 1), a novel crowdsourcing visual analytics environment for sensemaking of text data. CRICTO is motivated by the assumption that the performance of identifying and forming hypotheses from a large amount of text data can be enhanced when a diverse group of people are bringing their cognitive abilities and individualized knowledge to the integrated sensemaking process. CRICTO distributes text documents to crowdworkers for an initial schematization pass, where they are asked to read the text and present the information in a graph format that more concisely expresses the content in a form that can be readily used to build hypotheses [17]. This more structured representation also better supports computation, allowing the system to construct a cohesive representation by integrating the results from all of the crowdworkers' schematization microtasks.

The primary contribution of this work is new crowdsourcing techniques and workflow for visual text analytics, whereby a large pool of crowdworkers and one or more analysts contribute to generating a cohesive hypothesis from documents. In our presented workflow, a document dataset is split into a smaller set of documents, which are then assigned to different crowdworkers. Once workers utilize the visualization and analysis tools to make connections among entities (real-world physical and abstract objects) across the documents, more experienced analysts then complete the overall hypotheses-generation process based on schemata created by a large number of crowdworkers.

2 RELATED WORK

We drew inspiration for the design of CRICTO from two primary areas: (a) crowdsourcing analysis and data processing, and (b) visual analytics tools for sensemaking.

2.1 Crowdsourcing Analysis and Data Processing

There has been increasing interest in developing new tools and techniques to support shared analysis [13, 14, 15]. There are two main approaches evident in the research.

The first approach is fully collaborative, where many people interact and share their findings, collectively exploring and interpreting the data. Examples of this approach are Sense.us [18] and Many Eyes [16], which provide tools for sharing visualizations

*e-mail: {haeyong.chung|sd0051|sn0026}@uah.edu
†e-mail: candrews@middlebury.edu

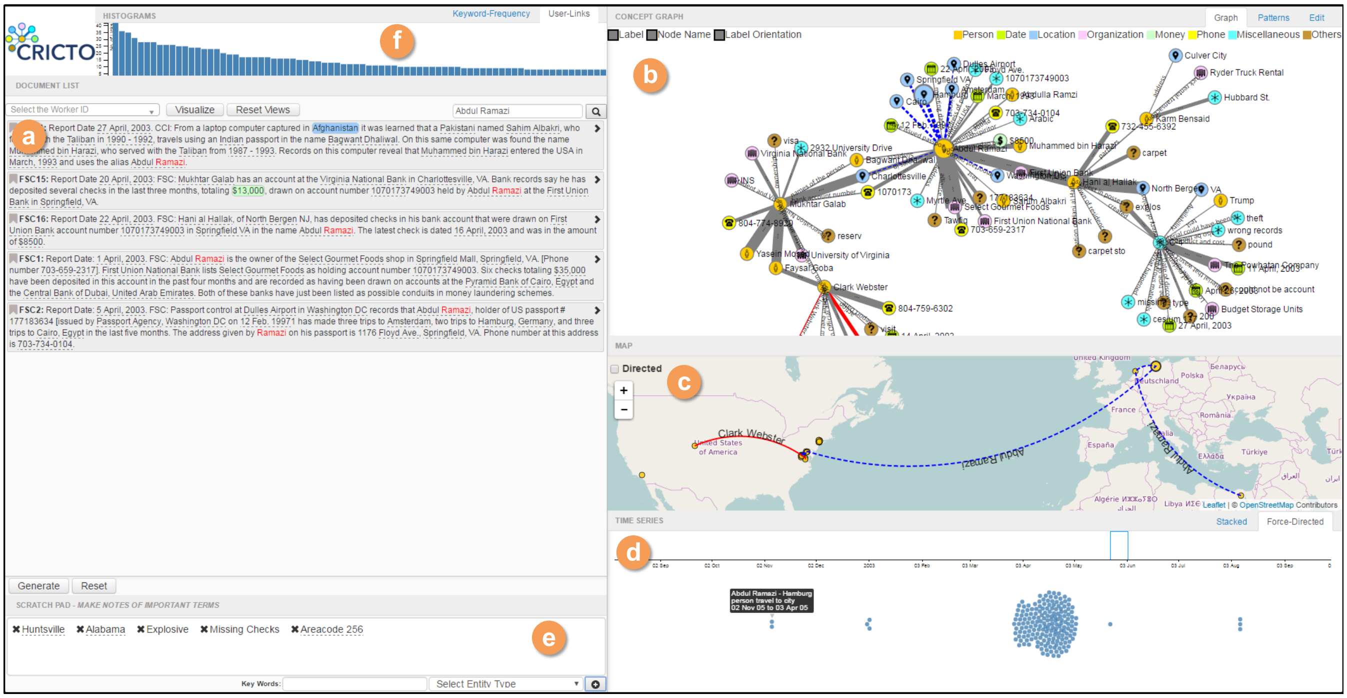

Fig. 1. CRICTO (CRowdsourcing Information sChematization TOol): a crowdsourcing visual analytics environment for sensemaking of text data. (a) Document view. (b) Graph view. (c) Map view. (d) Timeline view. (e) Scratch pad. (f) Bar chart (the number of links created by workers, etc.).

and data, and the use of message boards for discussion. Willett et al. further developed these concepts to enable collaborating users to identify and tag data in visualizations that either support or refute a hypothesis [19].

In contrast to CRICTO, in these systems, all participants have access to all of the data and visualizations. The benefit comes from more diverse points of view coming together to uncover different points of interest or interpretations for the same data. Interestingly, CRICTO somewhat inverts the collaborative visualization tools like Many Eyes, in that the product of the process is a visualization to which each participant has contributed some small piece.

The second approach is more about process and efficiency than collaboration, although many points of view still play a role. Amazon Mechanical Turk allows researchers and analysts to farm out low-level data processing tasks such as data annotation, classification, summarization, and editing to generate more meaningful information from data sources, which can then be evaluated by more skilled analysts [20]. The chief advantage of this approach is efficiency—inexperienced crowdworkers can participate by performing relatively short and simple tasks, after which the analyst can focus on the bigger picture by assessing the aggregate data. For example, Sorokin et al. [21] provided a data annotation framework that enables crowdworkers to annotate a number of images in a relatively short period of time. Kong et al. [22] also presented a crowdsourcing approach that facilitates the creation of relational linkages between text phases and the chart components to enhance data processing. Bernstein et al.'s Soylent explored crowdsourcing data editing and manipulation in word processing tasks, such as on-demand proof reading and the shortening of documents [26].

Several existing research studies in crowdsourcing suggest mechanisms for allowing a large number of human workers to cluster and classify data. For example, Luther et al.'s Crowdlines [23] is closely related to CRICTO in that it uses crowdsourcing to help users connect and organize information from various web documents. Willett et al. [14] made use of human workers to cluster different interpretations of charts via an interactive color-coding interface. The use of crowdsourcing to produce categories of responses to various questions was also introduced in Cascade [24]. This system uses crowdsourcing approaches to generate taxonomies of responses to questions posted on Quora.com. Verroios et al. [25] developed a new workflow to handle crowdsourced summarization tasks. Their approach is based on context trees, which provide a hierarchical workflow. Workers create summaries of small passages, but these are then hierarchically assessed and combined to create an overall summary. Kittur et al. [15] described the workflow of a crowdsourcing task that consists of dividing work into microtasks that are performed by individual workers, who then merge results to complete the required tasks.

CRICTO extends these crowdsourced analysis and data processing approaches with visual user interfaces and an integrated sensemaking workflow. In CRICTO, individual documents are sent out to crowdworkers for very basic data extraction tasks (i.e., the creation of entity links).

2.2 Visual Analytics for Sensemaking

An important step in analyzing a collection of documents is identifying relationships and performing link analysis [27]. CRICTO was inspired by a number of existing sensemaking tools designed to assist the analyst in performing the visual analytics task. For instance, Jigsaw [11] supports a number of ways to make visual connections between automatically-extracted entities in multiple documents. Additionally, Kang et al. [28] conducted an observational user study using Jigsaw and described how showing connections between different types of entities in Jigsaw was helpful for the sensemaking process in uncovering an embedded threat. However, Jigsaw's entity links rely exclusively on the co-occurrence of entities on documents—rather than taking into account the deeper semantic meaning of entities. Entity Workspace [29, 30] is a visual analytics tool that allows collaborating analysts to identify and connect entities. The system builds a graph to represent the collection of relationships, but the graph is not presented directly to the analysts; instead, it is used to drive recommendations and support memory. CRICTO's visualization views extend this prior research by employing similar visual

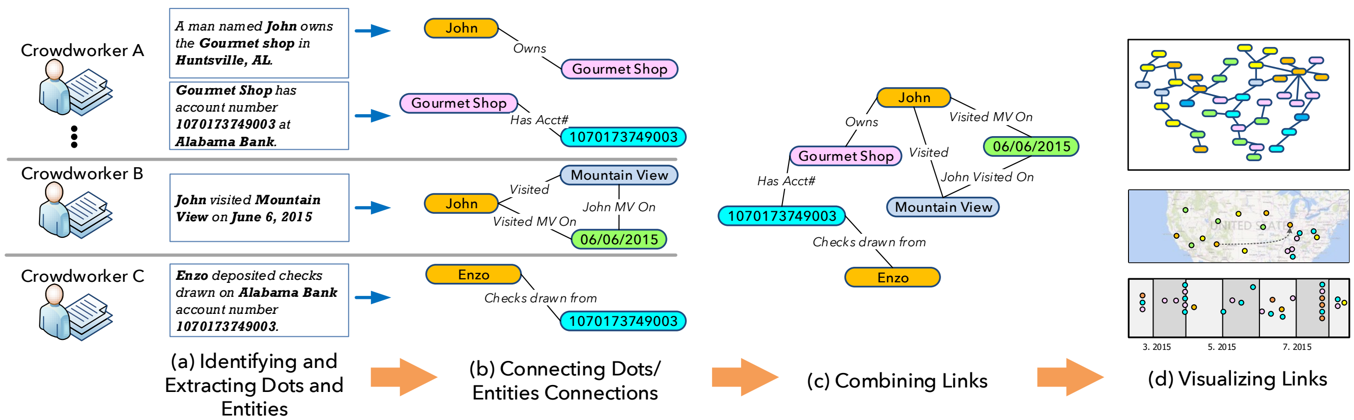

Fig. 2. Crowdsourced Schematization Process.

representation strategies (node-link graphs and multiple view visualizations) in order to help analysts uncover embedded plots in document datasets. However, CRICTO's visualizations were also designed to facilitate each crowdworker's schematization task.

Much existing research on collaborative visual analytics has focuses on merging and sharing individually developed graph views (based mostly on entities and links) to form aggregated analytic contributions that support the exploration of data for enhanced understanding. For example, CLIP [31] is a collaborative sensemaking tool based on entities and relationships; however, this tool emphasizes Linked Common Work (LCW), allowing collaborators to be aware of similarities to other collaborators' findings by discovering and linking their work visually. Brennan et al. [32] presented a framework for distributed collaborative visual analytics. In their framework, an individual user can develop a graph through independent analysis utilizing different graph view perspectives. Later, these individualized perspectives can be merged algorithmically with other user views to support collaborative analysis. The primary inspiration for CRICTO's visual analytics features is our prior experience developing VizCept [33], in which each user's individual analysis results contribute to creating a shared concept map. Specifically, in VizCept, each user can create a concept map from an independent line of investigation on the individual workspace, after which each individual concept map can then be merged with the shared concept map.

These existing visual analytics systems focus primarily on supporting collaborative awareness through the merged graph views. This model allows a group of workers to refer to or employ other workers' findings in real-time in order to assess, modify, or enhance their current work. In contrast, in CRICTO's workflow the final results from each worker's schematization task are integrated at a later stage of sensemaking by analysts. By not allowing the crowdworkers to see what others are doing, our goal is to reduce confirmation bias by soliciting purely independent viewpoints, and to encourage ownership over the process of completing the micro tasks to reduce "social loafing" [34].

3 Design Considerations

In developing CRICTO, our focus was on supporting the analysis of large collections of unstructured text. We wanted to make use of crowdsourcing to improve the sensemaking process. Superficially, this seems similar to the text summarization tasks done by Verroios et al. [25] and NLP-assisted text annotation tools such as [35, 36], but the challenge and goal are somewhat different. To this end, we made two primary design choices.

3.1 Crowdsourcing Schematization

We chose to try and break the problem down into microtasks that could be easily performed by crowdworkers. In order to maximize the size of the pool that could perform these tasks, and thus maximize the utility of using crowdworkers, we set three goals for the microtasks:

- Someone without experience with data analysis should be able to perform the required tasks
- Tasks should be simple and straightforward
- Workers should be required to commit only a small amount of time to undertake a task (e.g., less than 30 minutes)

The purpose of these goals was to maximize the size of the pool that could perform the simple microtasks and to reduce errors. Working on complex microtasks may increase crowdworkers' reluctance to engage with the task and/or invite errors—the results of which could be speculative or irrelevant explanations that contribute little to solving the actual problem [13].

The primary challenge of analyzing large text corpora is that semantic information and connections are buried in the data in a form that is very difficult to automatically extract. In order to manage the complexity of the task, one of the steps an analyst performs is *schematization* [17, 37]. Information is distilled, combined, and re-represented into structures (either internally or externally) that encode the knowledge the analyst has extracted from the data. These mid-level structures, or schemata, can then be pieced together to form hypotheses. This sub-process of transforming the raw text into coherent structures seems an obvious candidate for parallelization. Specifically, we decided to focus on extracting relationships between entities. This is a task that does not require training, nor does it require the crowdworker to have the context of the entire document collection. It is mechanical, but still a task that more easily and reliably done by a human.

We chose to use a workflow that consists of crowdworker and analyst phases in the sensemaking process. In the first phase, crowdworkers work competitively, working independently on small portions of the problem. In the second phase, a more trained analyst then merges the results and uses them as a platform to jumpstart the process of synthesis to gain understanding of the collection as a whole. This crowdsourcing workflow is based on a competitive model [8] which intentionally prevents workers from referring to each other's work. This model helps to avoid group-think issues, but also makes work parallelization easier [38].

3.2 Entity-link Representations

We posit the entity-link diagram as an appropriate structure to hold the results of the crowdworkers' efforts. Entity-link diagrams are a simple and straightforward encoding of relationships. Non-expert users can quickly grasp their meaning and can create them easily by linking two entities together with a simple relationship. The use of entity-link diagrams also helps to alleviate the problem of trying to combine the diverse findings of a large number of crowdworkers into cohesive insights. With CRICTO, each crowdworker can create simple link representations between entities across the documents. Such link representations facilitate the combination of different workers' results, eliminating the need for additional crowdsourcing sessions to combine them (links between entities), since the system automatically coalesces results from different workers by simply chaining the links they create, based on the entity.

Consider the following scenario. Suppose one report describes a suspicious person with the alias of "***John***" (Fig. 2a). One crowdworker may wonder if other documents could include the entity name John; and in searching, the crowdworker identifies two new relations: "***John*** *owns the* ***Gourmet Shop*** *in* ***Huntsville, AL***" and "*The* ***Gourmet Shop*** *is associated with the following international banking number:* ***1070173749003***." In parallel, suppose that a different crowdworker identifies a different report that states: "***Enzo*** *deposited a check drawn on* ***Alabama Bank*** *account number* ***1070173749003***." Even though those two crowdworkers did not directly identify an established relationship between Enzo and John, the combination of the links reveals the connection (Fig. 2b,c).

These crowdworker-selected entities have different semantic types (Person, Date, Location, Organization, Money, Phone, and Miscellaneous, etc.). Based on the semantic types of entities, the entity links can be categorized along the lines of What, Who, Where, and When, as described below.

- General/Social Network links (What and Who): This type of link comprises a range of abstract relationships across all types of entities. (e.g., social network, phone calls, money transactions, etc.)
- Geospatial links (Where): This type of link intends to represent the geospatial relationships of entities (e.g., the occurrence of an event, accident at a place, a person's home address, etc.)
- Temporal links (When): This type of link will indicate connections between any Time-based entity and other entities that explicitly represent a temporal relationship.

In order to better represent these types of links, CRICTO employs the Graph visualization, Map visualization, and Timeline visualization (Fig. 1b,c,d, and Fig. 2d).

Multiple entity-link categories can be also combined and represented in a visual representation. For instance, by combining the same link pattern of People-Location-Date entities from multiple workers, the analyst will be able to identify important trip routes for key players on the Map view (Fig. 3 & Fig. 1c).

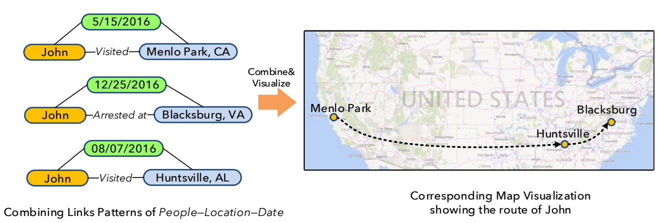

Fig. 3. Combining and representing multiple link categories.

4 THE CRICTO SYSTEM

CRICTO consists of four main views and additional tools (Fig. 1). The principal view (Document) displays the text documents, and the remaining three views (Graph, Map, and Timeline) offer various functionalities and analysis tools including visualizing, creating, and editing links created by a large pool of crowdworkers. All of these views are connected through brushing-and-linking.

4.1 The Document View

The Document view is designed to allow analysts and crowdworkers to search for, read, and begin to make sense of documents of interest from a large dataset (Fig. 1a). It also offers essential user interfaces for schematization tasks, whereby crowdworkers can create links between entities on documents along with short labels describing the relationships. To facilitate the process of labelling or annotating the important relationships between entities, the entities and their semantic types are automatically identified using LingPipe [39] and AlchemyAPI [40] prior to their presentation to the user and are underlined in the document. The entities are also color-coded for easy identification based on seven common entity types (see Section 4.2) as determined by the extraction algorithms. If the analyst believes a document to be important, she or he can also bookmark the document, which is shown on the additional division of the view.

The main purpose for the document view (aside from reading the contents) is to allow the crowdworker to identify relationships or links between entities. Links are created by clicking two entities consecutively (Fig. 4a,b), and then clicking the "generate" button (shown in Fig. 4c). The goal was to make this process as simple and low effort as possible.

Upon connecting two entities on one or more documents, the workers are asked to provide a link label for the entities' relationship and optionally to specify an event date or duration related to the relationship on the pop-up window (Fig. 4d,e). If the date is clearly specified on the document, they can also select the date as an entity and connect another entity to it.

Once two entities are linked on the Document view, the system instantly integrates and visualizes the links on the Graph view, the Map view, and the Time-line view.

Below the list of documents, we have provided a "Scratch pad", which can be used to create new entities not found in the documents. These new entities can be linked to existing entities using the process described above (Fig. 1e).

The Document view provides different levels of access to the full document collection depending on whether the user is a crowdworker or an analyst. Each crowdworker is assigned three primary documents that they are responsible for schematizing (see Section 5.2 for more details). Based on that person's initial schematization results, relevant documents are then introduced for further schematization. The analyst, on the other hand, has access to the full document collection and all of the entities and links that crowdworkers selected. A dropdown menu is provided to focus on the documents shown to any particular worker, and a search tool can be also used to explore the collection.

4.2 The Graph View

The Graph view displays all the collected entities and relationships discovered by the workers in a simple node-link representation (Fig. 1b). The entities are represented as circles using icons and colors to show type (Person, Date, Location, Organization, Money, Phone, and Miscellaneous, etc.), with the relationships shown as lines labelled with short descriptions. The graph uses a force-directed layout with pan and zoom navigation.

Multiple crowdworkers may create links between the same pair of entities. This apparently redundant labelling can be useful as it may indicate significance or the workers may have picked up on different aspects of the relationship and created different labels for the link. The Graph view indicates this with increased line thickness. If there are multiple labels for the same link, they are available in a tooltip.

Since a large pool of workers contributes to schematizing a large amount of text data and creating a global concept map together, viewing all the links created by all crowdworkers at the same time may be problematic and inefficient. Thus, the Graph view can visualize portions of the graphs based on the analyst's selection. First, the Graph view supports a drag-and-drop analysis from documents. For instance, if an analyst identifies a suspicious person to further investigate from related information while reading

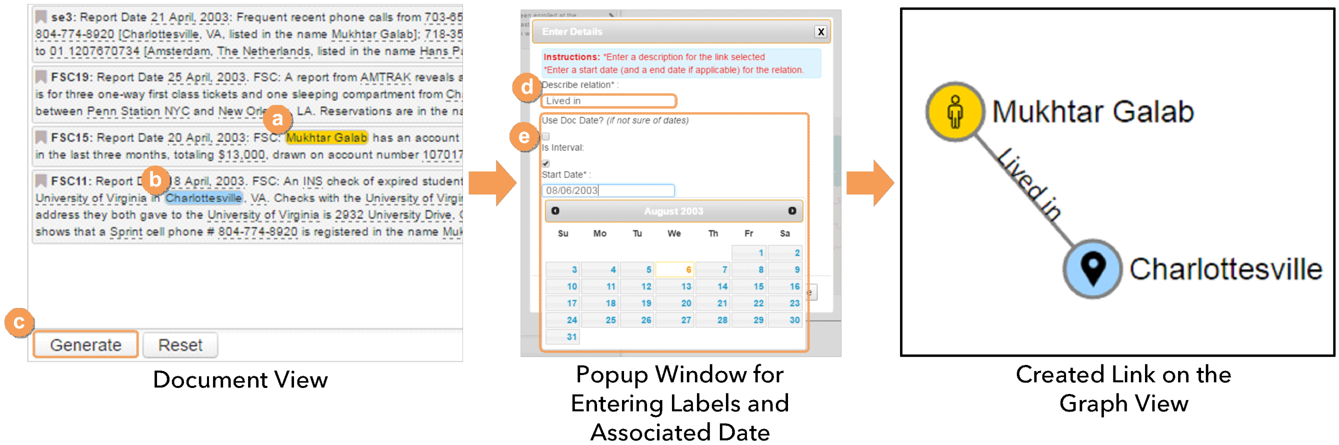

Fig. 4. Creating an entity link on the document view. The document view allows users to create links between entities by directly clicking on them. (a) People entity 'Mukhtar Galab.' (b) Location entity 'Charlottesville.' (c) Generate button for connecting two entities. (d) Input field for a link label. (e) Date selection UI for the link.

documents, he or she can examine the first level of connections to that person of interest by dragging and dropping an entity from the document onto the Graph view. This produces a graph centered on the person entity. The analyst can then analyze related documents and links for that person, and then merge and expand specific nodes directly on the graph. Analysts can also remove any unwanted links via a context menu. In addition, the analyst can select or filter a specific type of entity shown in the Graph view by using the check boxes for the entity types located at the top of the view. For example, the analyst might only wish to view people and phone entities on the Graph view. He or she can then apply both "Person" and "Phone" filters by clicking the associated legends on the Graph view (Fig. 1b, top).

4.3 The Map View

The Map view displays the geospatial relationships of entities connected by workers (Fig. 1c). In the Map view, an analyst or worker is able to identify and track the Location entities on a map. For example, when a crowdworker selects a location entity to create a link from the document (e.g., a link between "Virginia" and "George Washington," labelled with *birthplace*), the link is automatically visualized on the map. Each dot on the Map represents a link that has some connection with a specific location entity.

4.4 The Timeline View

The Timeline view focuses on visualizing the temporal aspects of each link. Workers can add a date to a link either when they first create the link by linking explicitly to a date entity, or later when they make connections between other entities (Fig. 1d). The Timeline view includes a horizontal axis that is used to show the dates and timescale. The events/links are represented by force-directed circles; while event duration is indicated through the use of rectangular bars parallel to the timescale that span from start date to end date. If analysts or workers hover over the circles with a mouse, the Time-line view provides details about the link such as associated entity nodes and labels, and the corresponding entities and links are highlighted in the Graph view.

4.5 Additional Analysis Tools & Views

In addition to the four main views, CRICTO provides a small collection of additional tools.

Bar Chart. The analyst has access to metadata about the efforts of the crowdworkers through a Bar Chart above the Document view (Fig. 1f). The visualization can show either the number of links created by each worker, sorted by count, or the number of times an entity was included in a relationship. This allows the analyst to look for important entities, or to evaluate the work of individual crowdworkers based on their productivity.

Edit View. Throughout the entire sensemaking process, all link data and properties can always be searched, added, edited, or removed to reflect updated findings via the Edit view. The Edit view allows analysts and workers to alter and refine each link element (the name of entity, link label, event date, etc.) in a spread-sheet format. If any single link is updated, the updates are immediately reflected on all of the visualization views.

Pattern Tool. The Pattern tool is designed to support more complex exploration of the data. The analyst can select up to three different entities or entity types. CRICTO will attempt to match the pattern and highlight connections among matching entities. For example, the analyst might specify connections between person and location entity types in a different color and line shape, and the Graph and Map view will then highlight all of the connections between people and location entities (blue dotted edges in Fig. 1b,c).

4.6 Implementation

To support the distribution to crowdworkers, CRICTO uses a client/server architecture. All the communication between clients and the server is in the form of Ajax (XMLHttpRequest) requests using JavaScript Object Notation (JSON) strings.

The client component provides visual interfaces for documents and visualizations. The user interface is implemented using Bootstrap and JQuery, while the visualization views are implemented with D3.js. We use Leaflet [41] to provide the Map view. The server, on the other hand, is implemented with Node.js backed by a MongoDB database. It is responsible for fetching and sending link and document data from the database according to the Client's requests, and for performing tasks such as link aggregation. Additionally, the Natural API [42] is used to perform the tf-idf weighting (Section 5.2) and entity normalization (Section 5.4), and OpenStreetMaps is used for geocoding location names to provide latitude and longitude data for the Map view.

5 Crowdsourcing Sensemaking Workflow

In this section we will discuss the CRICTO workflow in greater detail. The workflow consists of five phases, which are accomplished by the CRICTO system, crowdworkers, and analysts.

5.1 Pre-processing the documents

Before distributing the documents among crowdworkers, we run named entity extraction algorithms (LingPipe and Alchemy API) on the entire document dataset and store the extracted entities in the

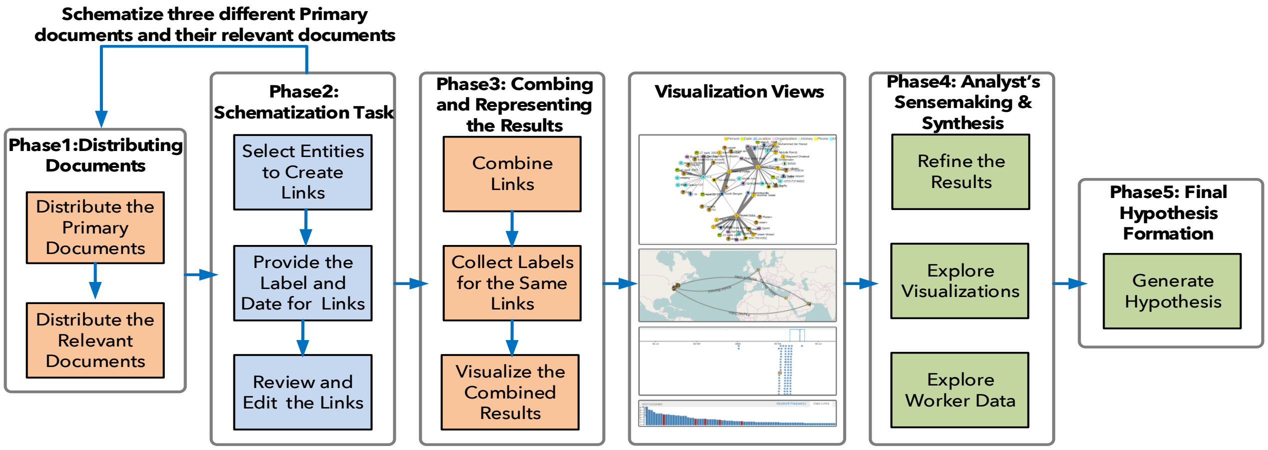

Fig. 5. CRICTO workflow for crowdsourcing sensemaking. Orange indicates CRICTO's task, Blue indicates crowdworker's task, and Green indicates analyst's task.

database. The documents in the database are also sorted according to their publication date and assigned a unique id number.

5.2 Dividing and Distributing Documents

The first phase of the workflow entails dividing and distributing documents across the different workers for the schematization task.

S1. Distribute the primary document: When a crowdworker logs into CRICTO, the system provides the worker with a primary source document, which is selected sequentially from documents in the collection. Once all the documents from the database are assigned as primary documents, the system iteratively re-assigns the first document in sequence to the next crowdworker.

S2. Distribute relevant documents: Once the crowdworker has created several links in the primary document (5 links are suggested as the minimum), a set of related documents are then assigned based on the created entities and links. We used a similarity metric based on the tf-idf scores of the chosen keywords [42, 43]. This metric determines the importance of selected entities in a document, assigning higher scores to documents in which the entities appear more frequently. According to this similarity metric, CRICTO selects the three highest scoring documents and provides them to the worker for further investigation. It should be noted that these relevant documents are selected from the same collection as the primary document, and can be assigned as a primary document to other crowdworkers as well.

S3. Repeat the process: Once the schematization for the relevant documents is concluded, steps S1 and S2 are repeated two more times, resulting in each worker working on 12 documents.

This phase ensures that crowdworkers are able to schematize a diverse set of documents, thereby increasing the likelihood of identifying important information from more documents rather than spending a great deal of time searching and foraging for documents manually (Fig. 5, Phase 1).

5.3 Crowdworkers' Schematization Tasks

Crowdworkers conduct schematization tasks for a set of documents by connecting entities in a document (or across multiple documents) and providing an annotation label for the relationship between the selected entities. Specifically, this process consists of three distinct steps (Fig. 5, Phase 2):

S1. Select entities to create links: Workers can select meaningful keywords and keyword strings by clicking on them while they read the presented documents (Fig. 4a,b). Alternatively, workers can also create their own entities manually using the Scratch pad rather than selecting existing entities on a document.

S2. Provide the date and labels for the links: Upon clicking the "Generate" button, CRICTO opens a dialog box where the worker enters a descriptive label and a date for the relationship between the entities. Optionally, the worker can specify that the date should be an interval and enter an end date (Fig. 4c).

S3. Review: Once the link has been added, it appears in all the visualization views (Fig. 5, Visualization views).

5.4 Combining and Representing the Results

After a crowdworker completes his or her work, CRICTO automatically coalesces the results into the visualization views (Fig. 5, Phase 3 & Visualization views).

The first part of this process is the normalization of entity names. Entities taken from different locations or those created by the workers refer to the same entity, but they may not be exact matches. To address this, CRICTO employs both stemming and lemmatization algorithms (according to the analyst's selection) to reduce the keywords to their root keyword, thereby facilitating the integration of those kind of entities. For example, in our live deployment, even though different crowdworkers selected two different entities—such as "chinchilla," and "chinchillas," —the system automatically recognized the root word ("chinchilla") and used that as the same entity for combining the links. It must be noted, however, that the stemming algorithm is not able to distinguish between entities that have some irregular word forms (e.g., mice and mouse, and feet and foot). In such cases, users can utilize the lemmatization algorithm to combine links with such entities based on dictionary look-up.

5.5 Sensemaking and Synthesis

Once results from all workers have been collected and combined by the CRICTO system, they will be available to the analyst (Fig. 5, Phase 4&5) for higher-level sensemaking. Typically, the analyst will begin the sensemaking task by reviewing the Bar Chart (Fig. 1f), either looking at the results of more productive workers or starting with entities with a large number of links.

It is expected that the analysts will make some refinement to the combined graph. Depending on individual skill level and personal investment in the task, crowdworkers may or may not select entities and create links carefully. Hence, combined results will likely include valuable links and labels, as well as considerable "noise"—i.e., information not directly related to a hypothesis or is faulty and unusable. Thus, it is crucial that analysts can identify and access the essential links more efficiently, while discarding any noise. For iterative refinements, the analyst can remove unwanted links and modify other link elements with the Edit tool (see Section 4.5 for details), while the analysis is progressing. Specifically, the analyst can change all information associated with a link, such as the link labels, the link date, linked entity names, etc. using the Edit tool.

Otherwise, it is expected that the analyst will explore the data using the collection of visualization views to build an

Dataset	#Documents	#Workers	Avg. Task Time	#Schematized Documents by Workers	#Total Links	#Entities Clicked	#Entities Typed	Duration (weeks)	Valid Links	Invalid Links
Small	41	90	11m45s	41/41	1330	2651	33	2	984	346
Large	1474	110	16m20s	476/1474	1710	3337	81	2	1200	510

Table 1. Deployment results of the small and large datasets.

understanding of any underlying stories, just as he or she might in a tool like Jigsaw or VizCept. The work of the crowdworkers becomes a foundation, which can be explored and refined, in forming a (or more) final hypothesis.

6 Evaluation

To understand and measure the effectiveness of CRICTO and its crowdsourced schematization for sensemaking, we employed both qualitative and quantitative analysis methods. We ran the evaluation in two different phases. First, we deployed CRICTO on Amazon Mechanical Turk, based on our suggested workflow, and conducted subsequent qualitative analysis of the crowdsourced output from the deployment. Second, we conducted a controlled lab study to determine whether the crowdsourced output from CRICTO was effective and efficient for sensemaking of document data.

6.1 Deployment

The deployment was executed in two phases based on the workflow described in Section 5. (1) For the first phase, we conducted deployments of CRICTO on Amazon Mechanical Turk for two document datasets. (2) During the second phase, three authors analyzed the crowdsourced output using CRICTO.

6.1.1 Deployment Design

Datasets. The deployment utilized (a) a small document dataset (the "Sign of the Crescent" dataset [44]), which has been used with students in academic settings; and (b) a significantly larger dataset (the "VAST Challenge 2007" dataset [45]). See Table 1 for details of both datasets. The small dataset contained 41 documents and featured 3 terrorist plots that needed to be uncovered. The large dataset, on the other hand, contained 1,474 documents and featured 5 plots related to a multinational animal trafficking racket.

Participants. Crowdworkers were recruited remotely through the HIT (Human Intelligence Task) website [12]. We posted only one HIT for each dataset at a time for two weeks. In the end, a total of 200 crowdworkers participated in our deployment (90 crowdworkers for the small dataset and a total of 110 workers for the large dataset). We paid each worker $0.75 per HIT, based on creating a minimum of 8 links per microtask. In total, we paid $67.50 for work performed on the small dataset, and $82.20 for work on the large dataset.

Microtasks & Procedure. All of the deployment and analysis procedures were designed according to the suggested workflow. The crowdworkers began the task by logging into CRICTO through links provided on the HIT website. When the workers logged into our tool, they were asked to review instructions and accompanying illustrative figures about how to use our system and tasks prior to proceeding to the actual task.

Data Collection & Analysis. During the deployments of CIRCTO, the activity logs and metadata, as well as links, labels, and their associated documents, were stored in the database. Based on a grounded theory approach [46], three of the authors engaged in an open-coding session during which each of the authors independently examined worker-created links and labels. This session resulted in the development of areas of concern or interest; the authors then collated this information through discussion and consensus to formulate the final coding rubric. The three authors then revisited the crowdsourced results and consolidated the final coding results by re-checking evidence in the associated links/labels using CRICTO. The authors then discussed each finding and reached a consensus on a selected subset of themes and patterns.

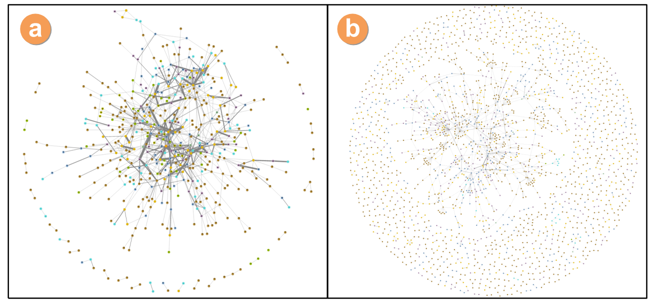

Fig. 6. Aggregated links from the deployment. (a) The small dataset: the large graph cluster in the middle represents the plots in the small dataset. (b) The large dataset: the largest graph cluster in the middle involved one of the five plots in the large dataset. Since all documents were not schematized in two weeks, the results included a large number of unmatched links which are appeared to be related to the other plots.

6.1.2 Results

Fig. 6 and Table 1 show the aggregated results of crowdsourced schematization tasks during the deployments of two datasets.

Small Dataset. By the end of the deployment of CRICTO on Amazon Mechanical Turk for the small dataset (41 documents), 90 workers created a total of 1,330 links (Fig. 6a) from their schematization tasks for the small dataset.

All 41 documents from the small dataset were schematized by crowdworkers. Overall, 2,651 entities (98.9%) were selected (clicked on documents) by crowdworkers; 33 entities (1.2%) were created via the Scratch pad. Many of the 33 custom entities consisted of multiple words (or a mixture of words and numbers) such as '50 Cal ammunition', 'land mines', 'Columbus Police', etc. Generally, these entities could not have been extracted as a single entity by either LingPipe or AlchemiAPI.

Among all worker-created links, 984 links (73.98%) could be categorized as *valid links* by the three authors in that the entities and labels of the links referred to meaningful information, and were labelled properly. On the other hand, 346 (26.1%) of the links were considered to be unequivocally *invalid* or trivial. These faulty links were created by workers who failed to select meaningful entities or establish a valid connection between entities (e.g., 'Dog'–'Pet', which they labelled *Dog is a Pet*.)

Large Dataset. With respect to findings connected with the use of the large dataset (1,474 documents), by the end of the deployment, the workers created a total of 1,710 links for two weeks. Similar to the result from the small dataset, a larger percentage (3,337 entities; 97.6 %) of entities are selected by workers by clicking on the documents, and 81 entities for links were created by manually typing them in.

Over the course of the two-week deployment period, 110 workers schematized a total of 476 documents from a total of 1,474 documents. 1,200 links (70.6%) were identified as *valid*. As indicated in Fig 6b, the largest graph cluster in the middle involved one of the five plots in the large dataset. Since 68% of the large dataset was not schematized within the two weeks, the results of the deployment included many unmatched links, which seemed to be related to the other plots (Fig. 6b). In addition, 520 links (29.8%) were invalid or faulty.

6.1.3 Analysis

We analyzed the crowdworker-created links and labels in order to determine how these links could be potentially different from algorithmically generated graphs, as well as what kind of unique knowledge could be extracted from both datasets by crowdworkers. Accordingly, a selected subset of our analysis results follows.

Complementing the sense of relationships. We assessed how multiple labels allow analysts to illuminate various aspects of the same link or entity, thereby providing complementary—or contradictory—interpretations of a relationship.

Overall, 547 (41.1%) links for the small dataset included more than two labels for the same relationship between entities. Frequently, several workers assigned labels with similar meanings for a specific relationship. For example, one of the links indicated that 'Bhagwant Dhaliwal' was employed by 'Empire State Vending Machines.' To express this relationship, different workers created multiple labels for this link—*Employed at, Employee of, works for, employee*—all conveying the same information. While such labels led to redundancy in describing a relationship for the same link, they could also point to stronger interest and importance for multiple workers.

However, in some cases workers created several annotated labels of different meaning describing the same entity link. These multiple meanings for the same entity-relationships were very useful in illuminating complementary aspects of the relationships and disambiguating relationships between two entities. For example, two workers created the same link between a person entity and 'Canada,' but assigned different labels for the relationship between those two entities. One worker assigned a simple fact that police in Canada were looking for him (*wanted by police*), but another worker indicated the reason that police were searching for him (*overstayed his travel visa*). In short, both links were created from the same sentence on the same document—but the two individual workers weighted different dots in the same sentence.

Also, there were links between the same two entities that were created from different documents. For example, there were multiple links between two person entities. A link from the first document revealed them to be room-mates; while the second label from another document showed them making reservations on Amtrak. The first label illustrates the relationship between the two persons, and the second label indicated a plan of a suspicious activity involving an Amtrak train.

Stimulating a new line of inquiry and questions. We observed that several link labels created by human workers were able to produce profound information. For such labels, multiple dots on a document were synthesized with each worker's domain knowledge and questions or hypotheses. As a result, analysts were able to identify the following types of crowdworker-generated labels, which could be useful in stimulating new lines of investigation during sensemaking tasks.

Since crowdworkers display a diversity of demographic backgrounds and domain knowledge, they will inevitably have knowledge of, and experience with, different fields, which is likely to contribute to greater thoroughness in identifying and annotating links. In this respect, some of the labels were based on workers' personal knowledge of a certain area and object of interest. For example, one of the workers described 'Alabama' as a *Religious region* based on her or his personal knowledge or belief, and another worker labelled a link between 'C-4' and 'Cesium 134' as *dangerous materials* based on her personal knowledge.

Importantly, several workers created labels that went beyond the relationships strictly found in the text. For example, a link between 'Fayasal Goba' and 'Northern Bergen' described the speculated relationship as *a possible location of the suspect*. Also, there were several links that were questions, such as *Did Clark Webster ever meet the forger?* ('Clark Webster'–'Muhammad Shamzi'), and *Is there a link between Joseph Nizar and this phone number in TEX?* ('Joseph Nizar'–'713-556-9213'). This type of label may help the analyst think about the data from an entirely new perspective.

Generating synthesized links. Several crowdworker-generated links did not represent simple extractions from the data, but instead resulted from some internal synthesis that linked several relationships, possibly across multiple documents. Specifically, we noted that 66 links in the small dataset and 111 links in the large dataset were created to connect entities found in different documents. For example, a single link between 'Mark Davis' and '718-352-8479' was synthesized from two different documents—the first contained both the address and 'Mark Davis,' and the other provided the identical address and the associate phone number '718-352-8479.' In other words, despite the fact that Mark's phone number was not specifically listed in the first document, in viewing the two documents the workers were able to identify that both the address of 'Mark Davis' and '718-352-8479' were connected through the same address—thereby enabling them to create the link representing a Person–Phone relationship.

6.2 Quantitative User Study

We conducted a controlled lab study to determine whether the output of crowdsourced schematization was both efficient and effective with respect to overall sensemaking task performance.

6.2.1 Method

In this study, we measured performance effects of crowd support (the output of our deployment) between two groups—one supported with crowdsourced links obtained from the deployment, and the other without crowd support. In our deployments and qualitative analysis, we envisioned that CRICTO would enable a large pool of crowdworkers to structure and extract valuable information from text documents, which each participant could then use to better identify and understand hidden plots embedded in the document datasets. Accordingly, we formulated the following two hypotheses for this quantitative study:

H1. *Crowd support leads to better overall sensemaking performance*.

H2. *Crowd support has a positive impact on sensemaking tasks and the use of analysis tools*.

Participants. We recruited 16 participants (aged 22 to 34, 14 males and 2 females) from a local university. The participants were PhD and MS students from the fields of computer science, business administration, electrical engineering and mechanical engineering. Each participant verbally expressed confidence in his or her ability to solve text analytics problems. Among the total participant cohort, 9 had prior experience with visualization tools and data analysis. The participants were randomized and divided into two groups (8 subjects each): 1) the non-crowdsourced (NCR) group was not able to access the crowdsourced links/labels from the two-week deployment; 2) the crowdsourced (CR) group could use the output of the crowdsourced schematization.

Task and Dataset. Each participant individually conducted a sensemaking task with the CRICTO system. We carefully replicated proven sensemaking tasks described in prior literature reports [9, 10, 47, 48] for CRICTO. The task was representative of a typical sensemaking task involving visualizations, where the participants had to read a series of documents carefully and use CRICTO's visualization tools to gain a global understanding of the data. Specifically, the users were asked to identify and synthesize clues spread across multiple documents and identify fictitious terrorist plots. We employed only the small dataset, since we

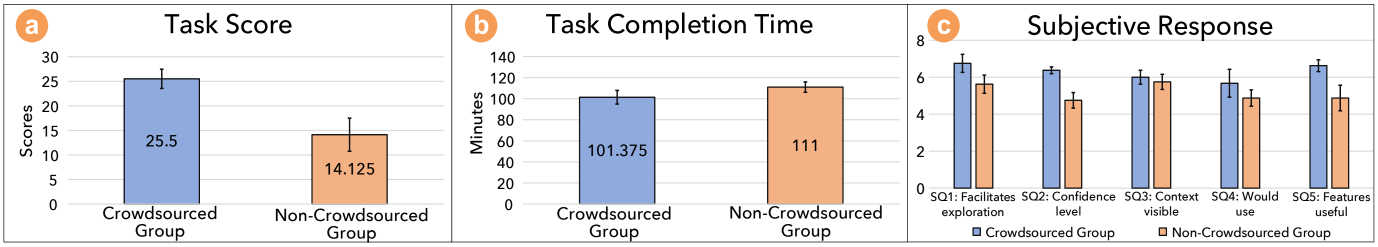

Fig. 7. Quantitative user study results. Average values are shown and error bars indicate the standard error. (a) Average scores of participants' analysis tasks for the two groups. (b) Average task completion times for the two groups. (c) Subjective response questionnaire results (SQ1 to SQ5) on a 7-point Likert scale (7=strongly agree, 1=Strongly Disagree).

needed to utilize a manageable-sized dataset that would enable participants to complete their analytic tasks within the two-hour timeframe of this controlled assay.

Procedures. The study sessions were undertaken using a Laptop PC equipped with Core i7-4510 CPU and 8GB RAM, which was connected to 24" monitor with 1920 x 1080 pixels. Prior to the onset of the exercise, participants were given a 20-minute tutorial on how to use the CRICTO system; they then had an additional 10 minutes to familiarize themselves with CRICTO's features with a test dataset. After the completion of the tutorial and the sample test, participants engaged in the actual experimental session of identifying fictitious terrorist plots and subplots of the small dataset using the CRICTO system. When participants started the study session, we provided them with a performance questionnaire (PQ1 to PQ4) that contained queries regarding the terror plots embedded in the small datasets. Although we gave participants a maximum of two hours to analyze the dataset, they were allowed to end the session if they were satisfied with their sensemaking tasks.

Data Collection and Analysis. For each study session, one of the authors observed and took notes. After the study session, we collected data from the exit survey and semi-structured interviews. All of the interaction logs and metadata, as well as updated links and labels created by each participant, were stored in the CRICTO's database. Specifically, we measured the analysis accuracy, the task completion time, and the subjective responses between the CR and NCR groups. Without knowing the subjective groups, we separately evaluated each participant's answer sheet from the performance questionnaire, and then discussed each response together in order to assign a final numerical score for subsequent tabulation.

6.2.2 Results

We evaluated two performance measures (the score of the performance questionnaire and task completion time) and subjective questionnaires (utilizing a 7-point Likert scale). Performance results are shown in Fig. 7a,b and subjective questionnaire results are illustrated in Fig. 7c.

Performance: We evaluated task performance based on the correct answers (a total of 54 points) to four main questions (utilizing the official VAST Challenge answer sheet [45]): PQ1) who are the players engaged in questionable activities? PQ2) what events occurred during a specific timeframe that are most relevant to the plot(s)? PQ3) what locations are most relevant to the plot(s)? PQ4) how are the plot(s) and subplots(s) and all findings from PQ1 to PQ3 tied to the plot and related to each other? For PQ1 to PQ3, we assigned 1 point to each correct 'people', 'event', and 'location' relevant to plots (a total of 24 points), and for PQ4, we assigned 10 points to each plot (i.e., a total of 30 points for 3 plots).

Based on the answers given by the participants in the two groups, we noted a statistically significant difference in the average task score between the crowdsourced (CR) group and the non-crowdsourced (NCR) group; both groups passed Levene's test for homogeneity of variances, and the normality of distributions was tested and confirmed by Shapiro-Wilk Test. An independent samples t-test was conducted on data with 95% confidence interval to compare the scores for the two groups. We found that there was a significant difference ($t(14)=2.908$, $p=0.011$) between the scores of the CR group ($M=25.50$, $SD=9.562$) and NCR group ($M=14.13$, $SD=5.566$) (See Fig. 7a). This result suggests that crowdsourced information did improve the analysis accuracy of CR participants, as compared with NCR participants.

In terms of task completion time (Fig. 7b), the sensemaking task was terminated when each participant determined that all the plots or answers were identified and there was nothing more to add to analysis results. While on average the CR participants (*M=101.38mins*, $SD=18.377$) completed their sensemaking tasks a little earlier than NCR participants (*M=111.00mins*, $SD=13.867$), the result of the Mann-Whitney U test indicated that there was no significant differences between the groups ($U=21.50$, $p=0.239$).

Subjective Response: The results for subjective responses are illustrated in Fig. 7c. After the sensemaking session, each participant completed a subjective response survey (SQ1 to SQ5) using a 7-point Likert scale (1 = strongly disagree, 7=strongly agree), in which both the CR and NCR participants rated their overall experience with the sensemaking task using CRICTO. On average, CR participants responded to all the questions more positively for both the use of CRICTO and crowdsourced results in comparison to the NCR group (Fig. 7c).

A Mann-Whitney U test was conducted on the subjective responses for both groups at the .05 level, indicating statistically significant differences in terms of confidence in analysis results and CRICTO's analysis features. Specifically, the crowdsourced data made CR participants more confident (SQ2) ($U=10.00$, $p=0.011$) in their analysis results ($M=6.375$, $SD=0.4841$) in comparison to the NCR group ($M=4.75$ $SD=1.479$). Also, the CR group ($M=6.63$, $SD=0.518$) found CRICTO's analysis features to be more useful (SQ5) ($U=11.50$, $p=0.021$) as compared to the NCR group ($M=4.88$, $SD=1.959$).

However, there was no significant difference between the two groups for the other three subjective response questions: CRICTO simplified exploration and analysis of the document dataset (SQ1) ($U=16.00$, $p=0.062$); the visibility of background information (SQ3) ($U=29.00$, $p=0.746$); and CRICTO for everyday use (e.g., study and summarization of text) (SQ4) ($U=19.00$, $p=0.162$).

7 Discussion & Future Work

Our user study results partially support our first hypothesis: **Crowd support leads to better overall sensemaking performance**. The CR participants who were aided by crowd support scored significantly higher (77.84% higher) in uncovering fictional terrorist plots embedded in the small data than the NCR group without crowd support. As we analyzed crowdsourced results, it was clear that output from the crowdsourced schematization did not merely represent extractions from the documents, but rather resulted from crowdworkers' synthesis efforts that linked important relationships across multiple documents. The crowdsourced visualizations also provided complementary details for the same links (see Section 6.1.3 for more details). This additional information encouraged CR participants to focus on high level synthesis. Additionally, all the CR participants identified at least one of the three plots in the small dataset (PQ4). In contrast, just one NCR participant identified one plot, while the other 7 NCR participants failed to identify a single plot within the two-hour time limit of the exercise.

Our results also partially support our second hypothesis: **CRICTO with crowd support has a positive impact on analysts' sensemaking tasks and the use of CRICTO's features.** We were able to verify this hypothesis based upon results of subjective responses and interviews with both groups, which was also partially evidenced by significant differences in the average Likert scores of two subjective response questions between the two groups. The crowdsourced results appear to increase the CR participants' confidence (SQ2) in the analysis results and make them feel CRICTO's features are more useful (SQ5) when compared to the NCR group.

Nonetheless, our evaluation results also served to reveal several limitations, as well as suggest potentially fruitful avenues for future work. Based on these observations, we discuss the identified problems and challenges in CRICTO's crowdsourced sensemaking environment.

Generalizability. For the studies described in this paper, we used data from intelligence analysis exercises. These are useful for evaluations, because they contain some fixed ground truth, which can be used to assess results. However, there are obvious problems with using crowdworkers to evaluate covert material.

Leaving aside the application to open source intelligence tasks [49], which would alleviate the concerns about sharing covert intelligence, CRICTO is well suited to other investigative sensemaking tasks like journalism. Sifting through one of the massive data dumps coming out of organizations like WikiLeaks would be an obvious use case.

In terms of its generalizability to other types of data—although it is not easy to extract semantic entities automatically from non-textual data such as images, maps, audio files, and even visualizations—CRICTO nevertheless enables crowdworkers to create entities manually, after which the created entities can be linked by crowdworkers. Thus, CRICTO's target tasks and proposed schematization approaches can be generalized to a variety of types of potential application domains and media data.

We caution, however, that our crowdsourcing workflows cannot be readily generalizable to any of the non-textual data types in terms of data distribution and task partition among different crowdworkers. Different similarity measures would need to be implemented to handle each different kind of media.

Data Coverage. As our evaluation results have shown, the small dataset had better coverage (with respect to how many documents were read and how many were read by more than one person for the two weeks) than the large dataset. Within the two-week period of the deployment, all 41 documents in the small dataset were schematized by an average of 5 different workers; moreover, on average, 32 documents were processed by the crowdworkers each day during that same period. For the large dataset, 476 documents out of a total of 1,474 documents were schematized by different workers, and 23 documents were processed on each day during the two weeks.

It is difficult for us to determine the data coverage of CRICTO and our crowdsourcing approach based on the amount of work accomplished, since document coverage per day can be affected by several inherent parameters of Amazon MTurk. For example, the number of workers who participated in our schematization task might be directly determined by how many HITs (Human Intelligence Tasks) were posted on the MTurk website at a given time (we posted only one HIT at a time), the desired qualifications of participants (we had no qualification criteria), compensation levels (75 cents per HIT), and so forth.

Hence, one important research avenue for examining the scalability of CRICTO would be to explore sweet spots including the appropriate data size to be divided among the workers and the crowdsourcing parameters to maximize both the data coverage and sensemaking performance. In so doing, we would be able to measure the effects of different volumes of data given to a worker, as well as how to improve data distribution based on the size of each document and the available number of workers.

Visual Scalability. A further scalability issue with CRICTO is that the representation of the aggregate of all of the crowdworkers' efforts is a node link diagram. However, as the number of documents in a dataset increases, so will the size of the graph (as shown in Fig. 6), eventually becoming so unmanageable as to transition from being an aid to the analyst to a new problem, i.e., determining what is important in the graph. In practice, we observed no statistically significant effects of crowdsourced data on the average task completion time. We attribute this result to the problem of checking a higher number of the connected entities.

We attempt to alleviate this problem by providing graph filtering tools that allow the analyst to focus on sub-graphs of interest. However, despite this, this problem caused 3 CR participants to run out of time (4 NCR participants ran out of time) and thus, fruitful avenues for future work lie in perfecting the graph views or adding visual representations (e.g. edge bundling clustering, meta-graphs, etc.) to further alleviate the issues associated with large graphs.

Bad links. Another issue we observed in the studies was that 26.01% in the small dataset and 29.8% in the large dataset of the links created by the crowdworkers were invalid, irrelevant or poorly labeled. Three of the CR participants in our study appear to have been misled by these links, wasting time pursuing non-existent plots.

There are several directions that the next-generation CRICTO might employ to help reduce this problem. One direction would be to include a qualification/rating component in the workflow, requiring crowdworkers to first perform certain qualification tasks [50] before being allowed to begin working on a given schematization task. To deal with redundant links, we can apply NLP lexical databases such as WordNet [51] to classify and extract redundant labels that convey semantically similar meanings or specific meanings that are of interest to analysts in the visualization views. Alternatively, we can come up with a reduced vocabulary of relationships so that crowdworkers are selecting from a set rather than picking them organically.

Computation vs. crowdworkers. With CRICTO, we have demonstrated that basic schematizing performed by crowdworkers can be usefully employed by an analyst to better understand a large dataset. We would like to emphasize that the goal of this approach is to leverage human perspective and judgement, not to replace computational tasks which can scale better. As such, future work will investigate ways that more advanced text mining approaches such as relation extraction might be incorporated into the workflow, potentially changing the role of the crowdworkers from specifying relationships to evaluating the output of the automated techniques. The question of what should be automated and what should be the result of human analysis is central to visual analytics, and a future version of CRICTO should allow us to study this mix by varying the amount done by each and evaluating the resulting links.

Acknowledgements

This work was supported by New Faculty Research Award from University of Alabama in Huntsville.

REFERENCES

[1] H. M. Wallach, "Topic modeling: beyond bag-of-words," in *Proc. of the 23rd international conference on Machine learning*, pp. 977-984, 2006.

[2] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Lingvisticae Investigationes,* vol. 30, pp. 3-26, 2007.

[3] A. Bagga and B. Baldwin, "Entity-based cross-document coreferencing using the vector space model," in *Proc. of the 17th international conference on Computational linguistics*-Volume 1, pp. 79-85, 1998.

[4] M. Banko, O. Etzioni, and T. Center, "The Tradeoffs Between Open and Traditional Relation Extraction," in *Proc. of ACL*, pp. 28-36, 2008.

[5] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining text data*, pp. 415-463, 2012.

[6] G. Demartini, D. E. Difallah, and P. Cudré-Mauroux, "ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking," in *Proc. of the 21st international conference on World Wide Web*, pp. 469-478, 2012.

[7] M. Schmitz, R. Bart, S. Soderland, and O. Etzioni, "Open language learning for information extraction," in *Proc. of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning,* pp. 523-534, 2012.

[8] J. J. Thomas and K. A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics:* National Visualization and Analytics Center, 2005.

[9] C. Andrews, A. Endert, and C. North, "Space to think: large high-resolution displays for sensemaking," in *Proc. of ACM CHI,* pp. 55-64, 2010.

[10] A. Endert, P. Fiaux, and C. North, "Semantic interaction for visual text analytics," in *Proc. of ACM CHI*, pp. 473-482, 2012.

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

[12] Amazon. (2017). Amazon Mechanical Turk Website. Available: https://www.mturk.com/.

[13] W. Willett, J. Heer, and M. Agrawala, "Strategies for crowdsourcing social data analysis," in *Proc. of ACM CHI*, pp. 227-236, 2012.

[14] W. Willett, S. Ginosar, A. Steinitz, B. Hartmann, and M. Agrawala, "Identifying redundancy and exposing provenance in crowdsourced data analysis," *IEEE TVCG*, vol. 19, pp. 2198-2206, 2013.

[15] A. Kittur, B. Smus, S. Khamkar, and R. E. Kraut, "Crowdforge: Crowdsourcing complex work," in *Proc. of ACM UIST,* pp. 43-52, 2011.

[16] F. B. Viegas, M. Wattenberg, F. v. Ham, J. Kriss, and M. McKeon, "ManyEyes: a Site for Visualization at Internet Scale," *IEEE TVCG*, vol. 13, pp. 1121-1128, 2007.

[17] P. Pirolli and S. Card, "Sensemaking Processes of Intelligence Analysts and Possible Leverage Points as Identified Through Cognitive Task Analysis," in *Proc. of International Conference on Intelligence Analysis*, p. 6, 2005.

[18] J. Heer, F. B. Viégas, and M. Wattenberg, "Voyagers and voyeurs: supporting asynchronous collaborative information visualization," in *Proc. of ACM CHI*, pp. 1029-1038, 2007.

[19] W. Willett, J. Heer, J. Hellerstein, and M. Agrawala, "CommentSpace: structured support for collaborative visual analysis," in *Proc. of ACM CHI,* pp. 3131-3140, 2011.

[20] A. J. Quinn and B. B. Bederson, "Human computation: a survey and taxonomy of a growing field," in *Proc. of ACM CHI*, pp. 1403-1412, 2011.

[21] A. Sorokin and D. Forsyth, "Utility data annotation with amazon mechanical turk," in *Proc. of IEEE CVPR Workshops*, pp. 1-8, 2008.

[22] N. Kong, M. A. Hearst, and M. Agrawala, "Extracting references between text and charts via crowdsourcing," in *Proc. of ACM CHI*, pp. 31-40, 2014.

[23] K. Luther, N. Hahn, S. P. Dow, and A. Kittur, "Crowdlines: Supporting synthesis of diverse information sources through crowdsourced outlines," in *Proc. of Third AAAI Conference on Human Computation and Crowdsourcing,* 2015.

[24] L. B. Chilton, G. Little, D. Edge, D. S. Weld, and J. A. Landay, "Cascade: Crowdsourcing taxonomy creation," in *Proc. of ACM CHI,* pp. 1999-2008, 2013.

[25] V. Verroios and M. S. Bernstein, "Context trees: Crowdsourcing global understanding from local views," in *Proc. of Second AAAI Conference on Human Computation and Crowdsourcing*, 2014.

[26] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, et al., "Soylent: a word processor with a crowd inside," *Communications of the ACM*, vol. 58, pp. 85-94, 2015.

[27] G. Chin Jr, O. A. Kuchar, and K. E. Wolf, "Exploring the analytical processes of intelligence analysts," in Proc. of ACM CHI, pp. 11-20, 2009.

[28] Y.-a. Kang, C. Gorg, and J. Stasko, "Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study," in *Proc. of IEEE VAST*, pp. 139-146, 2009.

[29] E. A. Bier, S. K. Card, and J. W. Bodnar, "Entity-based collaboration tools for intelligence analysis," in *Proc. of IEEE VAST*, pp. 99-106, 2008.

[30] E. A. Bier, E. W. Ishak, and E. H. Chi, "Entity workspace: an evidence file that aids memory, inference, and reading," in *Proc. of the IEEE International Conference on Intelligence and Security Informatics (ISI 2006)*, pp. 466-472, 2006.

[31] N. Mahyar and M. Tory, "Supporting communication and coordination in collaborative sensemaking," *IEEE TVCG,* vol. 20, pp. 1633-1642, 2014.

[32] S. E. Brennan, K. Mueller, G. Zelinsky, I. Ramakrishnan, D. S. Warren, and A. Kaufman, "Toward a multi-analyst, collaborative framework for visual analytics," in *Proc. of IEEE VAST,* pp. 129-136, 2006.

[33] H. Chung, Y. Seungwon, N. Massjouni, C. Andrews, R. Kanna, and C. North, "VizCept: Supporting synchronous collaboration for constructing visualizations in intelligence analysis," in *Proc. of IEEE VAST,* pp. 107-114, 2010.

[34] A. D. Balakrishnan, S. R. Fussell, S. Kiesler, and A. Kittur, "Pitfalls of information access with visualizations in remote collaborative analysis," in *Proc. of ACM CSCW*, pp. 411-420, 2010.

[35] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, and J. i. Tsujii, "BRAT: a web-based tool for NLP-assisted text annotation," in *Proc. of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 102-107, 2012.

[36] S. M. Yimam, I. Gurevych, R. E. de Castilho, and C. Biemann, "WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations," in *Proc. of ACL*, pp. 1-6, 2013.

[37] S. K. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in information visualization: using vision to think*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999.

[38] J. Heer and M. Agrawala, "Design Considerations for Collaborative Visual Analytics," in *IEEE VAST*, pp. 171-178, 2007.

[39] B. Baldwin and B. Carpenter. (2003). Lingpipe. Available: http://alias-i.com/lingpipe/.

[40] IBM. AlchemiAPI. Available: http://www.alchemyapi.com/.

[41] V. Agafonkin. Leaflet: an open-source Javascript library for interactive maps. Available: http://leafletjs.com/.

[42] C. Umbel. Natural: general natural language facilities for node. Available: https://github.com/NaturalNode/natural/.

[43] J. Ramos, "Using tf-idf to determine word relevance in document queries," in *Proc. of the first instructional conference on machine learning*, pp. 133-142, 2003.

[44] F. Hughes and D. Schum, "*Discovery-proof-choice, the art and science of the process of intelligence analysis-preparing for the future*

of intelligence analysis," Washington, DC: Joint Military Intelligence College, 2003.

[45] C. Plaisant, G. Grinstein, J. Scholtz, M. Whiting, T. O'Connell, S. Laskowski, et al., "Evaluating visual analytics at the 2007 VAST symposium contest," *IEEE CG&A*, vol. 28, pp. 12-21, 2008.

[46] J. Corbin and A. Strauss, *Basics of qualitative research*: Sage, 2014.

[47] P. Isenberg, D. Fisher, S. A. Paul, M. R. Morris, K. Inkpen, and M. Czerwinski, "Co-Located Collaborative Visual Analytics around a Tabletop Display," *IEEE TVCG*, vol. 18, pp. 689-702, 2012.

[48] P. Hamilton and D. J. Wigdor, "Conductor: enabling and understanding cross-device interaction," in *Proc. of ACM CHI*, pp. 2773-2782, 2014.

[49] R. S. Friedman, "Open source intelligence," Parameters, vol. 28, p. 159, 1998.

[50] M. Allahbakhsh, B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, and S. Dustdar, "Quality control in crowdsourcing systems: Issues and directions," *IEEE Internet Computing*, vol. 17, pp. 76-81, 2013.

[51] D. Gildea and D. Jurafsky, "Automatic labeling of semantic roles," *Computational linguistics*, vol. 28, pp. 245-288, 2002.