Newdle: Interactive Visual Exploration of Large Online News Collections

Jing Yang, Dongning Luo, and Yujie Liu
Dept of Computer Science
University of North Carolina at Charlotte
jyang13, dluo2, yliu39@uncc.edu

Abstract—In this paper, we present a novel visual analytics system named Newdle with a focus on exploring large online news collections when the semantics of the individual news articles have already been tagged. Newdle automatically conducts clustering and relation analyses on news articles and builds visualizations and supports interactions upon these analyses. By providing a novel topic overview in which the semantics and temporal features of the significant article clusters in a large collection are intuitively displayed, Newdle allows users to grasp the content of the collection in a glance. Through the rich set of interactions and visualizations provided by Newdle, users can effectively conduct in-depth analyses on topics, tags, and articles of interest. We have implemented a fully working prototype of Newdle, using the online New York Times RSS feeds as its example data input. We present several case studies to illustrate the effectiveness and efficiency of Newdle.

Index Terms—Online News, RSS, visualizations, text analysis, wordle, graph, visual analytics.

I. INTRODUCTION

Large amounts of news articles are released on the internet by newspapers, television news channels, and independent journalists that went online. Online news is an important information source since it provides timely, ambient information to not only the masses, but also business and political policy makers, social scientists, and analysts in other application domains. Online news is also an overwhelming information source. For example, The New York Times (NYT) (http://www.nytimes.com) delivers thousands of news articles in less than one month.

With the extremely large volume of online news, there is an urgent need for tools that allow users to effectively and efficiently browse topics, detect temporal trends, and search news of interest from this data source. Toward this end, numerous text analysis approaches, such as automatic news summarization (see [6] for an overview), have been conducted to automatically extract valuable information from large online news collections. However, even after such analyses, the amount of information to be delivered to users is still huge. It is challenging to effectively and efficiently transfer the large amount of automatic analysis results to users.

To address this challenge, visual analytics approaches [10] that tightly integrate automatic analysis and interactive visual exploration are promising. Such approaches use interactive visual interfaces to facilitate analytical reasoning so that human perception abilities and domain knowledge can be exploited together with computational powers to facilitate the reasoning process. Although there exist a few online news readers that employ automatic text analysis and use visualizations as their main interfaces, much more efforts need to be made to promote visual analytics in the important application of online news exploration. We argue that the following challenges are yet to be addressed for effective and efficient online news visual analytics:

- To use intuitive visual metaphors: Most persons who read or analyze online news are not visualization experts. The visual metaphors used for online news visual analytics approaches must be intuitive to the novice users;
- To convey semantic information: Online news is distinct from numeric data in that they contain rich semantic information. Effective online news visual analytics approaches should be able to transfer a large amount of semantic information to users in an instance.
- To be scalable: A small number of news articles can be explored in a simple text interface without any automatic analyses. The value of visual analytics approaches should be reflected in its better capability to support users in exploring large online news collections through information-intensive displays and interactive explorations powered by automatic analyses.
- To support interactive, in-depth analyses: Visual analytics approaches should allow users to conduct a wide range of in-depth analyses on large online news collections according to their particular interests by tightly integrating automatic text analyses and interactive visual exploration.

In this paper, we propose a novel online news visual analytics system toward addressing the above challenges. It is named Newdle, which stands for News Wordles, since wordles [11] are its major visual metaphor. Because extensive research has been conducted on automatic entity extraction and document summarization, we focus on exploration of large online news collections with an assumption that the individual news articles have been perfectly described by their tags. This assumption is reasonable since many online news sources provide manually generated tags for their news articles in the RSS entries. For example, the online news articles provided by NYT are manually tagged by library scientists with tags drawn from a normalized indexing vocabulary of people, organizations, locations, and topic descriptors. It is not hard to extend our approach to visualize untagged news by integrating existing entity extraction and document summarization algorithms.
Newdle conducts automatic clustering and path analyses upon networks composed of articles and tags. It uses wordles [11], the tag-cloud like, space efficient, and attractive visualizations, to visually depict the semantic information of article clusters. It also uses line graphs overlay on the wordles to visually convey the temporal trends of the clusters. A rich set of visualizations and interactions are provided in Newdle to allow users to perform the following in-depth visual analyses on large online news collections:

- **Topic analysis:** In this paper we define a topic as a cluster of articles that are closely related in content. Newdle automatically detects significant topics from a news collection and allows users to visually browse their semantic information and temporal trends in an information rich overview (See Figures 1 and 2 for examples). Newdle also allow users to examine the topics in detail (see Figure 3) and investigate topics related to a topic of interest (see Figure 4).

- **Tag analysis:** Attractive visualizations are used to encourage users to explore the most significant tags in a news collection. The visualizations present the tags within their semantic context so that the users can learn much about the tags before conducting any interactions. With interactions, the users can investigate tags of interest in more depth. For example, they can search articles and topics by tags, search tags related to a tag of interest, and even compare two tags with regard to the articles and topics they are involved in. Figure 5 gives an example of searching articles by tags. Figure 6 gives an example of tag comparison.

- **Article analysis:** Newdle allows users to search articles of interest by tags and topics. It also allows users to do extended reading for articles of interest, in which the users can access not only articles that share tags with the focus articles, but also articles that are indirectly related to the focus articles. Figure 7 gives an example of extended reading.

We have implemented a fully working prototype of Newdle, using the online NYT RSS feeds (a widely used data format for providing users with frequently updated content) as the example data source. Our case studies illustrate that our prototype is fun to use, effective, and efficient.

We discuss the related work in Section II. In Section III, we provide an overview of Newdle. In Section IV, we introduce the data extraction process in Newdle. In Section V, we describe the construction and automatic analyses of the article and tag networks. In Section VI, we describe the visualizations and the interactions in Newdle. In Section VII, we present a few case studies. In Section VIII, we draw our conclusion and present the future work.

### II. RELATED WORK

Newdle is related to two categories of previous work. The first one is online news visual exploration and the second one is visual analytics of large document collections.

There exist multiple efforts on visually exploring online news. Among them, Newsmap (http://newsmap.jp/) uses a treemap to visually present the automatic news article clustering results from Google News (http://news.google.com/). Each cell in the treemap represents a news cluster and its size is determined by the amount of related articles in the cluster. Titles of articles are displayed within the cells. Google News Timeline (http://newstimeline.googlelabs.com/) is a web application that organizes search results chronologically. It allows users to view news and other data sources on a browsable, graphical timeline. Narratives [4] tracks the attention to a news article in the form of references in social media (such as weblogs). It uses line graphs to visualize ways that concepts have changed over time and explore how those concepts relate to each other. VisGets [3] uses coordinated geospatial, temporal, and topical views to allow users to interactively explore online news. The users can find topics of interest through dynamic search queries. FeedDemon [1] is commercial software that allows users to read news in a text interface, tag news of interest, watch for news of interest, search news, and share news of interest with friends. Newdle is different from the existing approaches in that it provides a rich set of interactions to support many in-depth analysis tasks based on an underlying article network and an article-tag network.

There exist many efforts on visually exploring large document collections. For example, IN-SPIRE [12] employs clustering analysis and depicts hot topics in large document collections using the mountain and galaxy metaphors. The semantic information of the hot topics is displayed as labels. ThemeRiver [5] uses a river metaphor to depict the thematic strengths of significant keywords in large document collections. TRIST [7] allows analysts to formulate, refine, organize, and execute queries over large document collections. Its multipane view allows users to examine search results from different perspectives, such as clustering, trend analysis, comparisons, and difference. Jigsaw [9] is a visual analytic system that provides multiple coordinated views of document entities with a special emphasis on visually illustrating connections between entities across the different documents. COPLINK [2] presents a hyperbolic tree view and a springembedder graph layout of relevant entities to help law enforcement officials extract information from police case reports and analyze criminal networks. Different from the above approaches, Newdle presents the major topics in a document collection and their temporal trend using wordles combined with temporal overlay lines. It supports a wide range of visual analytics tasks making use of the underlying article and article-tag networks without explicitly displaying the networks.

### III. SYSTEM OVERVIEW

Newdle is a C++ program using WXWidgets for its GUI and Mysql for its data storage. It contains three components, namely the data extraction component, the network construction and analysis component, and the visualization and interaction component. We briefly describe them as follows:

- **Data extraction component:** Newdle automatically fetches news article entries from the NYT RSS feeds through Google Reader (reader.google.com). Google
Reader provides a standardized XML format for the RSS and caches news article entries for up to 30 days. Newdle uses a XML parser to analyze the retrieved entries to get the tags, titles, time stamps, and summaries of the news articles and the hyperlinks through which the original news articles can be accessed. The information is stored in a Mysql database and used by the other two components.

- **Network construction and analysis component**: Newdle constructs an article network and a bipartite article and tag network (article-tag network in short) to support the overview first, zoom in for detail visual exploration paradigm proposed by Ben Shneiderman and a wide range of visual analytics tasks. Clustering and path analyses are conducted on the networks for detecting hot topics and exploring relations among tags and articles. Their results are used to construct visualizations in the visualization and interaction component.

- **Visualization and interaction component**: Multiple coordinated views are implemented in Newdle. The system starts from a category index in which users can set a time period and select a news category of interest to define a news collection. A topic overview (see Figure 1 for an example) is then displayed in which the hot topics in the collection are visually presented. The semantics of the topics are displayed in wordles and their temporal features are overlaid on the wordles as line graphs. Users can conduct interactions such as relevant tag selection, topic selection, and tag comparison on this view. Most of the above interactions will trigger a variation of the detail view. In the detail view, multiple topics are displayed in rows, where their wordles and html boxes displaying an index of their articles are displayed. The detail view is used to present the results of article search, tag comparison, extended reading, and topic investigation (see Figures 3, reftopicext, 5, 6, and 7 for examples). Finally, a html browsing window is opened to display an online news article when its title is clicked in the detail view.

We describe these components in detail in the following sections.

**IV. Data Extraction Component**

Newdle can populate its database manually or automatically. In the manual mode, Newdle updates the database by fetching the NYT news article entries from Google Reader upon the request of users. In the automatic mode, Newdle automatically updates the database in a time interval defined by the users. Newdle keeps track of the latest update time so that entries already stored in the database will not be retrieved for another time. Users can define a life threshold. News entries older than this threshold will be removed from the database. Users can visually explore any news articles whose entries are in the database through Newdle.

**V. Network Construction and Analysis Component**

Newdle constructs an article network and an article-tag network for the news articles under exploration. They explicitly define the relations among tags and articles. Automatic analyses are conducted on these networks for generating the visualizations and conducting the interactions.

- **A. Article Network**

The article network describes the relations between the articles in a news collection. Each article is a vertex in this network. There exists an undirected, unweighted edge between two articles if and only if the articles share more tags than a given threshold, named the article relation (AR) threshold. The edge indicates that the articles are directly related. We allow users to interactively change the AR threshold to define the relations more strictly or more loosely, which will finally lead to topics of finer or coarser granularities. Note that the relations can be measured in other ways. We choose this simple approach since this measure is easy to understand and it is easy for users to change the topic granularities by adjusting the AR threshold.

We conduct the topic analysis upon the article network. In particular, graph clustering is conducted on the network. Graph clustering is the task of grouping the vertices of the graph into clusters taking into consideration the edge structure of the graph in such a way that there should be many edges within each cluster and relatively few between the clusters [8]. There is a large body of literature on graph clustering algorithms. The particular method used for graph clustering is however not relevant to this paper. We use the clustering algorithm provided by igraph (http://igraph.sourceforge.net) but any scalable and effective graph clustering algorithms can be used.

An article cluster is considered a topic in Newdle. The semantics of a topic is defined as the most shared tags among the articles in the cluster. The temporal feature of a topic is defined as the number of articles in this cluster over time.

We further define the relevance between topic A and topic B as the number of edges that connect the vertices in topic A and the vertices in topic B. We also define the distance between article A and article B as the length of the shortest path between them in the article network. They are calculated upon requests from the interactions.

- **B. Article-Tag Network**

The article-tag network describes the relations between tags and articles in a news collection. It is a bipartite network in which an article is connected to each of its tags. Given a tag, its relevant articles are all the articles connected to it. To find tags that are relevant to a given tag, we first select all the articles connected to the given tag, and then find all the tags connected to them.

**VI. Visualization and Interaction Component**

In this section, we first discuss wordles, the major visual metaphor used in Newdle, and then present the visualizations and interactions in Newdle.
A. Wordles

The major visual metaphor used in Newdle is wordles [11]. Wordles are attractive, tag cloud like visualizations that pack a large number of tags with varying font sizes and colors into a small screen space. User studies conducted by Viegas et al. [11] suggested that wordles have emotional emphasis, colors, and layout that enhance the meaning. They are attention keeping visuals that catch the attention of people. They also encourage users to explore the most significant tags in them.

Wordles seem suitable visualizations for online news since they address our online news visualization challenges well: we need intuitive visual metaphors and wordles are easy to understand, fun to use, and attractive to users; we need to convey semantic information and wordles display tags; we need to display large amounts of information and wordles are space efficient; wordles attract user attention to the most significant tags and thus they help users perform visual analytics tasks such as finding significant tags.

In our approach strong semantic meanings are attached to the wordles. First, in the Newdle views, each wordle represents the semantics, namely the most shared tags, of an individual news topic. In such a wordle, the tags convey the when, where, who, why, and what information about the topic being represented so that users can learn them in a glance. The semantic meanings of the individual tags are clear in such a context. Using wordles in this way significantly distinguishes our approach from many existing approaches where tag clouds are used to display a set of not necessarily related tags.

Second, each tag is associated with a tag category, such as location, person, organization, and topic (the categorization is provided by NYT). We assign a unique color to all tags of the same category to distinguish tags of different categories and help user focus on tags of a category of interest. A qualitative color scheme provided by ColorBrewer (www.colorbrewer2.org) is used. As stated in ColorBrewer, qualitative schemes are best suited to representing nominal or categorical data and the hues are used to create the primary visual difference between classes. For example, in Figure 1, all location tags are in gold and all person tags are in purple. A user interested in location tags can thus look for gold tags in the wordles.

Third, the size of a tag in a wordle is proportional to the number of articles with this tag in the topic. Thus big tags, namely the significant tags in a topic, easily catch the eyes of users.

To reveal the temporal trends of a topic, we overlay a line graph on a wordle to reveal the daily number of articles published on this topic (see Figure 1 for an example). The line graphs in different wordles use the same scale to allow users to compare the topics with regard to their strengths over time. For example, we can immediately notice significant, bursty topics in Figure 1.

B. Visualizations

Newdle has three basic views, namely the category index, the topic overview, and the detail view.

Category Index: Users start their visual exploration in Newdle from the category index. In this view, they set up three important parameters for the following visual exploration, namely the time period, the news category, and the AR threshold.

The time period and the AR threshold can be set through the menu bar of the category index. The default time period is set to be from 30 days ago to present, during which the news entries are guaranteed to be accessible. Users can interactively change the time period. It can be within the default time period or beyond it if the data is available in the database.

The default AR threshold is set to be 2. In other words, articles sharing two or more tags are considered to be directly related. Users can change this threshold for changing the topic granularity. In the case study presented in Section VII, the topic granularity is interactively changed.

In the canvas of the category index, news categories that are available from NYT online RSS feeds, including world news, US news, business news, technology news, sports news, and health news, are listed. A wordle is displayed for each category to help users select a category of interest. It displays the most frequent tags in that category during the selected time period.

By clicking a category name, a news collection is defined and its topic overview will be displayed to the users. The users can always go back to the category index to change the three parameters and start a new exploration. In the following sections, when we talk about a news collection, we mean the collection defined by the current time period and the category of interest.

Topic Overview: The topic overview allows users to browse the significant topics in the news collection to learn their semantics and temporal features in a glance. Many interactive explorations, such as topic analysis and tag analysis, can be conducted in or triggered from this view.

To generate this view, the article network is first constructed upon the news collection using the given AR threshold. Clustering is conducted on the article network. The resulting article clusters are topics to be explored. We sort the topics by the number of articles in them in descending order. A wordle (see Section VI-A) is then generated for the biggest topics and displayed in the overview in a line by line manner following the order. Figure 1 shows an example topic overview. We will discuss the interactions upon this view in Section VI-C.

Detail View: The detail view has multiple variations according to the function it serves. The basic form of the detail view (see Figure 3 for an example) has multiple rows in it, each for a topic. The left side in a row is the wordle of the topic. The right side in a row is a html box in which the titles and tags of the news articles are displayed in a list. The colors of the titles range from white to blue to indicate their ages: the more bluish, the older the news article is. The order of the titles in the list is also related to their ages. By clicking the show detail tag in a html box, more information about the articles, such as their time stamps and summaries, will be displayed. A user can click a title to open a html browsing window in which the original news article is displayed.

The detail view has five variations, namely the detail view of topics, the search result view, the topic investigation view,
the tag comparison view, and the extended reading view. We will discuss them in Section VI-C.

C. Interactive Visual Analysis

Newdle supports users in conducting in-depth visual analyses on topics, tags, and individual news articles, which are introduced as follows.

1) Topic Analysis: Organizing news articles into topics allow users to explore a large news collection effectively and efficiently. Newdle provides the following interactions to allow users to analyze topics:

- **Topic browsing in the overview**: The significant topics are displayed in the topic overview as wordles (see Figure 1 for an example). The sizes of the wordles are fixed to make sure that the tags in them are readable. When there are more wordles than a screen can hold, a scrolling bar is provided so that users can navigate in a canvas larger than the screen space.

- **Topic browsing in the detail view**: In the detail view of topics (see Figure 3 for an example), the significant topics are displayed in the same order as in the topic overview, with their detail html boxes displayed besides the wordles. Users can browse the titles of news articles in the topics, access news articles of interest, and start extended reading from this view.

- **Topic reconstruction**: If users are unsatisfied with the granularity of the topics, they can adjust the AR threshold to reconstruct the topics. For example, if a user finds that many unrelated articles are included in the same topic, she can increase the AR threshold to construct topics with more coherent content. Figures 1 and 2 shows two topic overviews on the same news collection with different AR thresholds.

- **Topic investigation**: Users can examine a topic in detail and investigate its relevant topics in the topic investigation view. It is triggered by double clicking a topic of interest from the topic overview. Figure 4 shows an example. In this view, the topic of interest (TOI) is placed in the top row. The other rows display topics relevant to the TOI, sorted by their relevance to it. In their html boxes, there are star symbols in front of the titles. They indicate how the articles are related to the TOI: A 3 star symbol indicates that the article is directly connected to one or more articles in the TOI. A 2 star/1 star symbol indicates an indirect connection with a distance of 2/3 in the article network. The titles are first sorted by their stars and then by their ages.

2) Tag Analysis: Users can find significant tags by looking for big tags from the wordles in the topic overview. Beside this manual inspection approach, Newdle provides the following interactions for conducting in-depth tag analyses:

- **Selecting tags**: Users can first select a set of tags from the wordles and then conduct other operations upon the selection. Selected tags are highlighted by red surrounding boxes. A tag can be manually selected/unselected. In particular, when users hover over a tag in the wordle, a white surrounding box of the tag will be displayed. Users can then press the mouse button to change the selection status of the tag. Newdle can automatically select tags relevant to a set of tags of interest (KOs). To trigger this operation, users first add the KOs into the current selection and then click a relevant tag search button. To perform the selection, the system will first use the article-tag network to find articles with the KOs. For each article found, its tags in the wordle of the topic to which this article belongs will be added into the current selection. In Figure 2, the tags in yellow background were selected using a relevant tag search starting from the tag Haiti. Note that the same tag can be selected in multiple wordles. That a tag is selected in one wordle does not mean that it is also selected in other wordles. Users can clear the selection by clicking a button.

- **Searching articles by tags**: Users can click a search button to search articles with all or any tags in the current selection. The search results are displayed in the search result view, as shown in Figure 5. It displays the research results of the tag Haiti. In this view, each row is a topic containing one or more resulting articles, sorted by the number of search results they contain in descending order. Only resulting articles are displayed in the html boxes, with the tags searched highlighted by italic.

- **Marking tags**: Several buttons with different colors are provides for marking selected tags. By clicking such a button, the background color of the tags in the current selection will be changed to the color displayed on the button, namely that they are marked in that color. For example, the tags related to Haiti are marked in yellow background in Figure 2. Users can start a new search after they mark the tags in the current selection. Thus they can compare the results of multiple selections by marking their results in different colors. Marked tags can also serve for guiding users in finding topics of interest.

- **Comparing two tags**: When there are two tags selected, users can click the tag comparison button to trigger the tag comparison view to find the relation between the tags. Figure 6 shows the tag comparison view for tags Yemen and Abdulmatallah, Umar Farouk. On the top of the view, the two tags being compared are displayed. We call the tag on the left tag 1 and the other one tag 2. Articles with both tags or any of the tags are displayed in their topics. In the html boxes, there are two circles in front of each title. The left one is filled with color if the article has tag 1, and is empty if it does not have tag 1. The right one indicates whether the article has tag 2 in the same way.

3) Article Analysis: We allow users to conduct extended reading for an article of interest. By checking the button on the left of an article in a detail view, users can define the article as an Article Of Interest (AOI). They can then click the extended reading button to trigger the extended reading view for the selected article. Figure 7 shows an example of the extended reading view. The top row displays the AOI. The other rows display articles that are directly or indirectly connected to the AOI in the article network, organized by their topics. The star
symbols indicate the distance from an article to the AOl. A symbol with three stars/two stars/one star means a distance of 1/2/3.

VII. CASE STUDIES

We present a few case studies in this section to illustrate how Newdle can be used. The news collection under exploration is the NYT world news from 12/17/2009 to 1/19/2010. There are 917 news articles in this collection.

A. Haiti Earthquake

In this case, we start from the topic overview shown in Figure 1. The top right wordle catches our eyes since there is a big burst in its line graph. By reading its tags, we know that it is about the Haiti earthquake. We want to make the topic granularity finer and thus we change the AR threshold from 2 to 3. A new topic overview is generated, as shown in 2 without the yellow marks. We want to search and examine topics about Haiti from this view.

We start by highlighting the news topics related to Haiti. To do it, we click the tag Haiti in a wordle, conduct a relevant tag search, and mark the selected tags in yellow background. The resulting display is shown in Figure 2. We notice that there are three significant topics highlighted by the yellow background of their tags. The first one is about the disasters and emergencies in Haiti, the second one is about humanitarian aid to Haiti, and the third one is about the latest update about the Haiti earthquake. We then click these topics one by one to examine them in detail.

We can also use an alternative approach to perform the task. First, we select the tag Haiti from the topic overview and search articles with this tag. The search result view is shown in Figure 5. We find all the topics we noticed above. We also browse the news titles in the topics and click the titles of interest to read the articles in full detail.

B. Bombing Attempt on a Christmas Day flight to Detroit

In this case, we investigate the second to the left most topic in the top row in Figure 2, which seems to be breaking news according to its line graph.

We start by clicking its wordle from the topic overview to open a topic investigation view, which is shown in Figure 4. The topic of interest is in the top row. From the line graph in the wordle, we notice that it burst sometime after 12/17/2009. We drag the scrolling bar of its html box to examine the earliest news article in this topic. It was published on 12/26/2009. We click the title to access the original news article. It talks about a bombing attempt on a Christmas day flight to Detroit that was prevented by the passengers and crew members. We then examine more articles in this topic. Most of them are the follow-ups of this event and articles discussing security and warning systems and airports and airlines, as suggested by the wordle. After examining the topic itself, we browse the relevant topics presented under the top row in Figure 4. We find that the topic in the second row discusses Yemen, terrorism, and Al Qaeda and the topic in the third row was about the reactions of Obama Barack and United States to this bombing attempt.

We are curious why Yemen was related to the bombing attempt. To investigate this question, we select the tags Yemen and Abdalmutallab, Umar Farouk and compare the two tags. The latter is the name of a person (we learn it from its color) that appears in both topics. In the tag comparison view shown in Figure 6, several articles with both tags pop up. By reading the news with the title Yemen Says Bomb Suspect Met With Qaeda Figures, we learn that Abdalmutallab, Umar Farouk is the person who made the bombing attempt and he met with operatives of Al Qaeda in Yemen before setting out on his journey.

We can also investigate the event using an alternative approach. In this approach, we start from the detail view of the topics, as shown in Figure 3. We select the earliest news article in this topic, and conduct extended reading for this news article. The view is shown in Figure 7. It is interesting to see that the topic in the third row has been there before the bombing attempt from the line graphs. By examining the html box, we notice that there are some news articles in this topic published before 12/26/2009 that are indirectly related to the bombing attempt. They talk about Yemen and Al Qaeda.

VIII. CONCLUSION

In this paper, we present Newdle, a novel online news visual analytics system. Its visualizations and interactions are supported by article network analyses and article-tag network analyses. By employing the intuitive wordle metaphor to display semantic information with temporal information overlaid, the Newdle visualizations effectively transfer the semantics and temporal features of major news topics in a large news collection to users. The rich interactions and visualizations provided in Newdle support users in performing a wide variety of in-depth analysis tasks, as illustrated in our case studies.

It is important to point out that Newdle is not limited to exploring online news collections. Other document collections can also be explored in it. We plan to use Newdle in several other applications, such as science policy analysis and email visualization. We will also conduct formal user studies to evaluate the effectiveness and efficiency of Newdle in the future.

Acknowledgement. This material is based upon work supported by National Science Foundation under Grand Award Number IIS-0946400 and the U.S. Department of Homeland security under Grand Award Number 2008-ST-108-000002.

Disclaimer. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

REFERENCES

Fig. 1. The topic overview with a coarse granularity (the article relation threshold is 2). The biggest 12 topics are shown in the figure. More topics can be accessed using the scrolling bar on the right of the display.

Fig. 2. The topic overview with a fine granularity (the article relation threshold is 3). A relevant tag selection has been conducted for the tag Haiti and the selected tags are marked by yellow background. The marked tags highlight the topics about Haiti.
Fig. 3. The detail view of significant topics.

Fig. 4. The topic investigation view. The topic of focus is displayed in the top row.
Fig. 5. The search result view. It displays the articles with the tag Haiti.

Fig. 6. The tag comparison view.
Fig. 7. The extended reading view. The article of interest is displayed in the top row.


