Footprints: A Visual Search Tool that Supports Discovery and Coverage Tracking

Ellen Isaacs, Kelly Domico, Shane Ahern, Eugene Bart, and Mudita Singhal

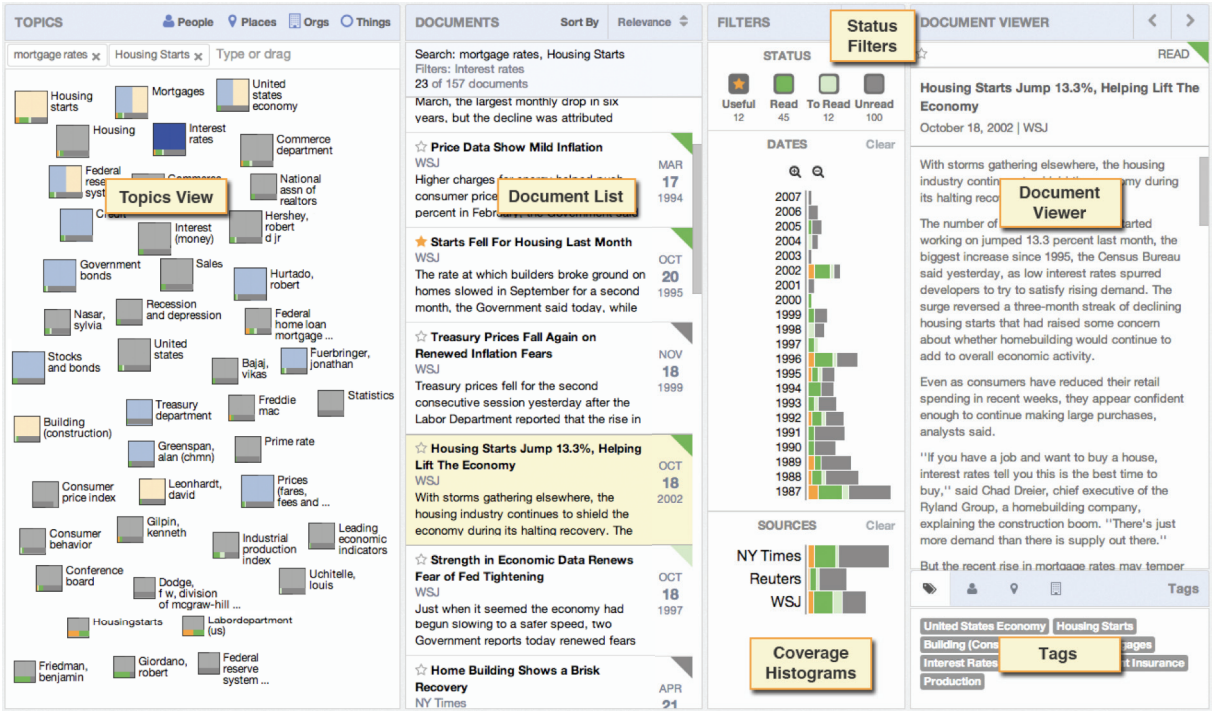

Fig 1. User Interface of Footprints, a topics-based document search tool. Footprints supports exploratory search by helping analysts (a) discover information that they may not know to look for and (b) keep track of their coverage along several dimensions, so they can avoid missing important information and know when to stop searching.

Abstract—Searching a large document collection to learn about a broad subject involves the iterative process of figuring out what to ask, filtering the results, identifying useful documents, and deciding when one has covered enough material to stop searching. We are calling this activity "discoverage," *discovery* of relevant material and tracking *coverage* of that material. We built a visual analytic tool called Footprints that uses multiple coordinated visualizations to help users navigate through the discoverage process. To support discovery, Footprints displays topics extracted from documents that provide an overview of the search space and are used to construct searches visuospatially. Footprints allows users to triage their search results by assigning a status to each document (To Read, Read, Useful), and those status markings are shown on interactive histograms depicting the user's coverage through the documents across dates, sources, and topics. Coverage histograms help users notice biases in their search and fill any gaps in their analytic process. To create Footprints, we used a highly iterative, user-centered approach in which we conducted many evaluations during both the design and implementation stages and continually modified the design in response to feedback.

Index Terms—discovery search visualization, visual cues, discoverage, coverage tracking, document triage, interactive histograms

1 Introduction

Current commercial search tools are breathtakingly good at helping people locate a specific piece of information, but they are less well suited to researching a broad topic where the challenge is in figuring out what to search for and knowing when one has found enough relevant material. This process of "discovery search" is different from conventional search, and neither Web-based search engines nor browsers provide good support for such exploration. As a result, people improvise, often opening up dozens of browser tabs to retain listings of search results and articles of interest. The result can be a row of indistinguishable tabs that do little to help people keep track of their search process. Further, current search interfaces do nothing to help people figure out how extensively they have read and whether they are missing anything important. In an open-ended discovery search process it is often hard to know when one has read enough and can stop searching.

With little assistance in managing the search process, people researching a broad topic are prone to narrowing their search scope too quickly in an effort to reduce the overwhelming number of search results to a manageable size [17, 18]. In doing so, intelligence analysts have been shown to inadvertently exclude key

The authors are with Palo Alto Research Center (PARC) and can be reached at firstname.lastname@parc.com.

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014. Date of publication 11 Aug. 2014; date of current version 9 Nov. 2014.
For information on obtaining reprints of this article, please send *e-mail to: tvcg@computer.org*.
Digital Object Identifier 10.1109/TVCG.2014.2346743

documents from consideration, leading to incomplete or even mistaken interpretations of the topic. The problem stems not just from difficulty in uncovering the relevant information but in lack of awareness that something has been missed. While much research has focused on helping people discover information [3, 5, 20, 25], there has been little work on showing people what they are missing.

Research into the search process has shown that discovery in a large document collection is typically an iterative process in which researchers enter query terms, examine the results, and modify the query terms until they are satisfied with the results [19, 23]. Pirolli and Card [19] characterize the research process as having two phases: a foraging loop and a sense-making loop, with lots of iteration within and between the two phases. In the foraging loop, analysts explore the document space and read documents to extract information, collecting a "shoe box" of evidence in the process. In the sense-making loop, they develop and evaluate a hypothesis of their analysis. The authors note that people are prone to many cognitive biases that can lead them to inaccurate or incomplete conclusions during the search process, even if they are motivated to avoid such errors. For example, working memory limitations can cause people to limit the amount of evidence they explore. People tend to rely on familiar sources and fit data into their existing belief structures. And once they formulate a theory, people tend to look for confirming evidence.

Visual cues and indicators can help analysts avoid such biases in the foraging loop, yet those cues are often either missing or inadequate in search tools. Existing commercial tools are aimed at efficiently presenting results, but they do not help users evaluate their *coverage* of those results. Chuang et al. [2] make a similar distinction, noting that models need to help analysts both make inferences about the underlying data (which they call *interpretation*) and evaluate the accuracy of those inferences (which they call *trust*). They claim that many systems for visualizing large document collections have problems with both interpretation and trust.

We took on the challenge of designing a tool that would enhance analysts' ability to not just discover and interpret information, but also visualize the extent of their investigation of the material. We are calling this process "discoverage," a combination of *discovery and coverage*. By discovery we mean figuring out what relevant information is available, and by coverage we mean investigating enough to develop a sufficiently complete understanding of the material. Good coverage requires tracking the extent of one's exposure to relevant information, identifying gaps in one's knowledge, and filling in those gaps. We took the position that, although we cannot directly evaluate the quality of researchers' analysis, we can provide "information scent" [1, 30] to help them notice potential biases and gaps in their coverage so they can correct and improve their analysis.

We developed a prototype system called Footprints meant to support discoverage. We designed it for a community of intelligence analysts who are regularly tasked with producing a summary report about a broad subject or question without knowing exactly what to search for, sometimes under tight deadlines. We were motivated to consider the coverage aspect of the research process because these analysts were particularly concerned about potentially missing critical but perhaps obscure information. One analyst explained their requirements succinctly when she said, *"We're terrified of missing things. We want to make sure that we're [searching] as broad as possible, but that we also focus very quickly on what the most relevant information is."*

Although Footprints was designed to address the needs of a particular set of analysts, we believe it addresses a broader need for tools that help people research a subject in a large document space without missing key information. To create Footprints, we used a highly iterative and user-centered process that incorporated detailed input from researchers and analysts at many points throughout the design and implementation phases.

2 Requirements Gathering

To understand the analysts' requirements, we participated in a two-day workshop that included a representative group of analysts, technologists tasked with supporting the analysts' needs, and social scientists who had conducted a study of the analysts' work habits. The outcome of this workshop confirms an extensive literature on the needs of analysts [6, 7, 8, 9, 15, 16, 26], so here we provide only a brief summary of these analysts' practices and the key design requirements that emerged from the workshop.

2.1 Analysts' Practices

The analysts are organized into teams that cover certain geographic areas or subject areas. They are discouraged from becoming highly specialized, so it is common for them to move from one subject area to another every couple of years. Most analysts spend the first part of their day monitoring traffic to see what is new in their area, and then they report their findings at a daily meeting. The manager filters the updates from his or her team and passes the most important information up the chain. The analysts are expected to be knowledgeable about the latest developments in their areas and so work hard to stay current.

The analysts' reports are typically quite short, about 1 page for a typical brief and 3-4 pages for a longer-term report. These reports may be written in response to a request, perhaps from a policymaker, or the analyst may pitch an idea for a report. They typically spend 2-3 days writing a report, but an urgent one might be requested hours before it is needed. A longer-term report may be in progress for 2-3 months, sometimes longer.

The analysts have access to a huge number of documents from a wide range of sources, including news articles, research reports, and internal reports and briefs. They cannot possibly read everything that has been written on a subject, and yet they are tasked with not just staying on top of ongoing events, but interpreting their significance and anticipating possible developments. Typical subjects for a report might be:

- "How likely is it that event X will come to pass, what would be the likely outcomes, and how would that affect our interests?"
- "Provide a profile of emerging public personality Y."
- "What is technology Z? What are its capabilities, why was it developed, and when might it get used?"

With these kind of open-ended questions it is difficult to know when the answer is complete. Given their current text-based search tools, the analysts said their typical strategy was to generate a Boolean search query and then try to get the results down to a manageable size without having to read or skim every document. They often reduced the list by adding "AND NOT" to their query. As one analyst explained, searching now is *"coming up with the longest list you can think of and figuring out where to put ANDs, ORs, and NOTs."*

The analysts are particularly concerned about falling prey to unconscious biases and yet they acknowledge that doing so is difficult to avoid. They want their tools to help them identify contrary evidence and recognize when they are neglecting certain information. Through a series of exercises, the workshop participants generated the following requirements for a tool that would support their work. They wanted it to help them:

1. Figure out what to search for in a large document collection
2. Understand the context of information
3. Characterize large issues quickly
4. Understand relative importance
5. Detect trends or emerging changes early
6. Notice the decline or absence of issues
7. Avoid missing important information
8. Notice contradictory evidence
9. Know what others are saying (experts, public figures, the public, etc.)
10. Understand the timeline of events

11. Understand the depth and breadth of their coverage, including any biases

Rather than designing a single visualization to support the analysts' task, we designed a tool that integrates multiple coordinated views to support the whole discovery search process – including discovery of information, document triage, efficient reading of documents, and assessing document coverage. Several other systems also combine multiple visualizations to let users to visualize and manipulate document sets in different ways [4, 14, 25, 32]. Footprints is novel in introducing the notion of coverage tracking and in providing a collection of cues to help the user find useful documents. That is, it not only provides cues about where to go, it also leaves a trail of where one has been, hence the name Footprints.

The next section describes the design of Footprints, followed by a description of its implementation and then the results of two evaluation phases.

3 Footprints Design

To help us design Footprints we came up with a hypothetical but representative question an analyst might be asked to research: "What are the main factors that led to the explosive growth in the U.S. housing market and that ultimately led to the economic crisis of 2008?" At a high level, Footprints' design consists of four components, as shown in Fig. 1:

- **Topics View:** Shows a subset of the topics extracted from the document set that are related to the search query, and provides a visual mechanism for generating and refining a search.
- **Coverage Histograms:** Shows the distribution of documents across date, source and coverage status, and allows analysts to filter documents across multiple dimensions.
- **Document List:** Lists the documents in the results set and offers the ability to triage documents.
- **Document Viewer:** Shows the content of the documents and provides options to tag and annotate documents.

These four components help analysts both discover what they need to know and track their coverage as they read. The Topics View and Coverage Histograms work together to visualize and support discovery, and the Document List and Coverage Histograms together provide tools to track coverage and to visualize and correct any gaps. The following sections describe the two key components, the Topics View and the Coverage Histograms, and how they support discoverage. A video of Footprints' main features is available at vimeo.com/98558826.

3.1 Topics View

Footprints supports the discovery of information primarily through the Topics View with support from the Coverage Histograms.

3.1.1 Visual Topics-based Search

The *Topics View*, shown in Fig. 2, is both a visualization of the topics extracted from a document collection and a visuospatial mechanism for *generating and refining* a search query. That is, analysts use the Topics View both to discover topics related to a keyword search and to create queries by dragging topics from the topics pane into the search box, either adding to an existing query or creating a new one.

Each box represents a topic, and its size indicates the number of documents in the full document set related to that topic. Along the top of the pane is the search box that can hold either keywords entered manually or topics dragged from below. To generate an initial query, the user types a search term into the search box, for example "mortgage rates" in Fig. 2. Based on a *Magnet Model*, topics related to that search term float up from the bottom of the pane to fill the view, with the topics most relevant being "pulled" up toward the top and those less related coming to rest lower down in the view. This type of physics-based model has been shown to be effective and natural for users to understand [21, 31]. It allows users to scan down the view from top to bottom to discover related topics that may also be of interest. The horizontal placement of topics did not map to a particular meaning because we planned to use that dimension for a 2D-search feature (discussed in 5.1.2), but we unfortunately did not have time to implement that feature in this version.

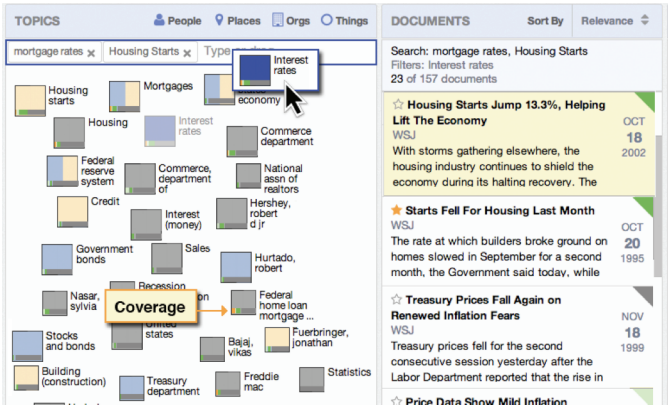

Fig 2. Topics View with a topic and a document selected. Light blue indicates a connection to the selected topic, orange shows a connection to the selected document, and topics with both colors are related to both.

Footprints offers a number of "information scent" cues that suggest places to look for relevant documents. First, instead of showing a graph of the entire topic space, the Topics View surfaces a subset of topics, few enough that all the labels can be shown so that analysts can comfortably scan them and select topics of interest. Second, it surfaces the topics most closely related to the search while also balancing coarse- and fine-grained topics. That is, topics that appear frequently are often broad (e.g. finance) and so less helpful for discovery of unanticipated topics than specific but rare topics (e.g. industrial production index). The Topics View shows more fine-grained topics than a simple relevance algorithm would.

Third, when an analyst selects a topic, other highly related topics become highlighted in light blue (see Fig. 2), drawing her attention to other topics that might also be of interest. Similarly, when a document is selected, topics related to it are highlighted in orange, suggesting other topics to consider if that document is informative. When both a document and a topic are selected, topics related to both are shown with a split blue-orange highlight, indicating that those topics might be especially fruitful to explore.

Finally, the analyst can refine her search by dragging additional topic boxes into the search box, defining a search *Concept*, or a set of topics and keywords. In Fig. 2, the user previously dragged the topic Housing Starts into the search box and is now dragging in Interest Rates to refine her search further. As additional topics are added to the search, the topics pane adjusts so that strongly related topics move up and weakly related ones move down; some new topics may emerge from the bottom and others may drift down and disappear. As the concept is more clearly defined, the proportion of fine-grained topics increases.

With each search, the Document List also adjusts to show the documents resulting from the query (Fig. 1). Each time the analyst selects one or more topics in the Topics View, the Document List is filtered to show only documents related to those selected topics. By selecting topics and documents, analysts can combine both top-down and bottom-up approaches to discovering information. That is, by getting an overview of relevant topics in the Topics View, they can identify documents that might be of interest (top-down), and by reading documents they can identify other related topics to explore further (bottom-up).

This design is meant to help analysts uncover unanticipated topics and documents as they explore a search space. After initiating a search with a keyword or phrase, an analyst can scan the Topics View to identify related topics. As she drags more topics into the search box to refine her search concept, increasingly relevant documents appear in the Document List and, more importantly, more fine-grained topics begin to emerge in the Topics View, helping her discover unanticipated topics to explore. Selecting those topics lets the analyst filter the Document List to reveal other useful documents she may not have found. With those documents selected, the Topics View reveals related topics, potentially suggesting yet new avenues of exploration. Through this iterative process of using topics to identify useful documents and documents to reveal new topics, analysts can identify relevant documents on topics that may be far afield from their original search terms – documents they would not have known to look for.

Other tools have used similar types of topic or word graphs, such as WordBridge [11], Phrase Nets [27], Text Tree [13], Many Eyes [28], and TopicNets [5], but those systems aim to show the whole space. Our focus is on helping the user figure out where to look by surfacing only the most relevant topics, balancing the presence of coarse and fine topics, offering visual cues of relatedness to both topics and documents, and making it easy to adjust the topic space by dragging topics into the search box.

3.1.2 Topics Filtering

Footprints also helps analysts narrow in on certain *types* of topics – people, places, organizations, and things – through the *Topics Filter* (Fig 3). To filter the view, the analyst simply selects the associated attribute and the topic boxes that don't match fade away (but do not disappear), making those of interest pop out visually.

Fig 3. Topics View with filtered to show only People and Orgs.

3.2 Coverage Histograms

The Topics View is meant to support the foraging phase of exploratory search by suggesting query terms the analysts might not have anticipated. The Coverage Histograms have a dual purpose in that they support both foraging and sense-making.

3.2.1 Document Discovery

The coverage histograms, shown in Fig. 4, are a visualization of the document set along two key dimensions, namely Date and Source. The Date histogram shows the number of documents according to the date they were published or released, and the Source histogram shows which organization published or generated the document. The Date can be set to cover different time ranges, and it can be zoomed in and out so that each bar represents a year, a month, a day, or even an hour, which might be useful if a news event is breaking. The Source histogram can also be modified to show individual sources (e.g., articles from the New York Times, Wall Street Journal, Reuters), as shown in Fig. 4, or scaled up to types of sources (e.g., news articles, briefs, scholarly papers, classified reports), and the analyst can choose which sources or types of sources to display.

By visualizing the distribution of documents along these two dimensions, the histograms reveal an initial sketch of the subject matter, specifically the timeline and who was writing about it. For example, in Fig. 4 it is easy to see that there was a lot of coverage in the late 1980s to the mid-1990s and then it died down, with an uptick in 2002. The New York Times covered it more extensively than the other sources did. Seeing this distribution, an analyst might decide to focus on 1987 through 1996 and check what happened in 2002. She would do so by selecting one or more bars at a time to filter the set of documents by those attributes. In addition, the histograms make it easy to generate compound queries that combine dates and sources, for example, documents published from 2000 to 2002 by the New York Times or Reuters.

Overlaid on these histograms are indicators of the analysts' coverage through the document set. This feature works together with the *Document Triage* mechanism provided in the Document List, described next.

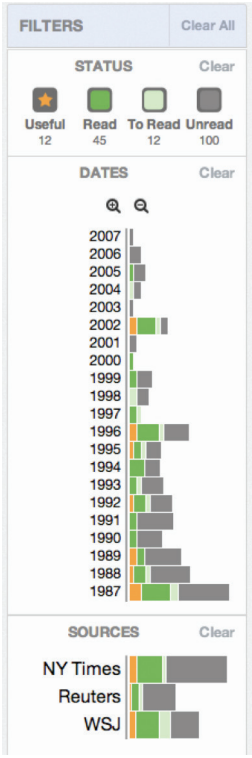

Fig 4. Coverage Histograms show the distribution of dates and sources in the document set, overlaid with the user's coverage status.

3.2.2 Document Triage

The Document Triage mechanism helps analysts read more efficiently by allowing them to first skim document headers, mark those that look promising, and then easily return to just the marked ones to read them all at once. Each document header has a *Status Marker* in the upper right corner (Fig. 5), which starts out grey, indicating the document is *Unread*. When the analyst identifies one she would like to read, she clicks the corner to change it to light green, marking it *To Read*. She might conduct several searches in

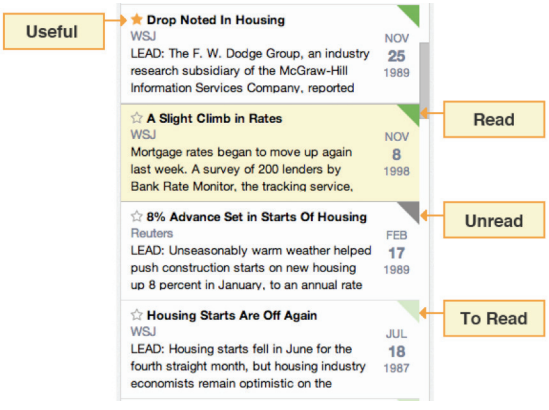

Fig 5. Document Triage. Users mark a document To Read by clicking on the corner and it turns light green. When they read (select) the document, it turns dark green. They can mark documents Useful by clicking on the star to the left of the title.

sequence, marking documents To Read each time. When she is ready to start reading, the analyst can click on the To Read button in the Status Filters area (Fig. 4) to display only To Read documents and efficiently read the most promising articles all in a row.

Once a document is read, its status marker changes to dark green. If the analyst finds a document useful, she can click the star to the left of the title. Later, when she is ready to develop her analysis, she can click on the Useful status filter to show only those documents. In this way, the analyst can quickly return to the key documents to construct her analysis of the subject.

3.2.3 Coverage

Coverage through a document set is shown by overlaying the four statuses on the Date and Source histograms and on the topic boxes in the Topics View. As shown in Fig. 4, the histogram bars are divided into sections, indicating the proportion of those documents that have been marked Useful (orange), were Read (dark green), or marked To Read (light green), with the remaining left as Unread (grey). Similarly, a horizontal strip along the bottom of each topic box shows the same bar, indicating the portion of documents on that *topic* with those statuses (see Fig 2).

These visualizations give analysts an understanding of what they have read, what they have missed, and what they have found most useful along these dimensions (date, source, and topic). Most importantly, these visualizations make it easy to notice any biases in their coverage and to counteract them. For example, if an analyst sees that she has not read many documents from 2005 she can click on that bar to filter the results and fill in her understanding of that year's events. She can get very specific in her query, for example showing just documents about adjustable mortgages (click on its topic box) published in June of 2008 (date bar) by Reuters (source bar) that she hasn't read yet (Unread filter). She can also run certain compound queries in one click by selecting a section of a histogram bar (e.g., the orange portion of the 2008 bar to see Useful documents published that year).

Document status attribute is saved across search queries, giving another coverage cue. As the analyst explores a search space, she can quickly find documents related to ones she previously marked useful by clicking on the Useful filter. She can also see whether each new search is revealing many new documents; once there is relatively little grey in the bars, she knows she has probably uncovered most of the relevant material.

Together, these histograms are meant to aid analysts who are concerned about missing important information. They allow analysts to see the attributes of the documents they have read and skipped, exposing any inadvertent gaps in their reading, and making it easy to quickly locate documents that fill in those gaps. Another benefit is that they help an analyst with limited time to research efficiently. She can choose where to focus her reading and be aware of the type of information she may be missing. The idea is that since one can't know everything, it is useful to be aware of what one doesn't know.

Interactive histograms have been used in other tools to show the distribution of certain attributes in a dataset [3, 10, 12, 14, 30]. Kwon [12] is especially relevant because that system also supported intelligence analysis and also overlaid another attribute on the bars. However, the overlaid attribute was a static aspect of the data rather than a dynamic variable reflecting the users' activity, in our case their progress through the dataset.

In summary, the topics view combined with the coverage histograms support exploratory search by helping users uncover topics they may not have anticipated, and by visualizing their coverage of the material across certain dimensions (topic, date, and source) as they read through documents. The histograms prominently display any gaps in their coverage, and make it easy to filter the document list so the user can fill those gaps or at least be aware of what they ignored. The visualizations also aid in suggesting topics, date ranges, and sources to explore.

4 Implementation

Footprints is a Web application implemented using open source web technologies. An Apache Tomcat server running a SOLR [33] engine sits at the back end and supports RESTful calls from a browser-based front-end that uses JQuery, D3, JSON, HTML5 and CSS for querying and rendering. It is built on the assumption that a useful set of topics can be extracted from a large document collection using one or more linguistic grammar-based techniques as well as statistical models. The tool is agnostic as to how topics are identified, but assumes that topic labels are descriptive and unique.

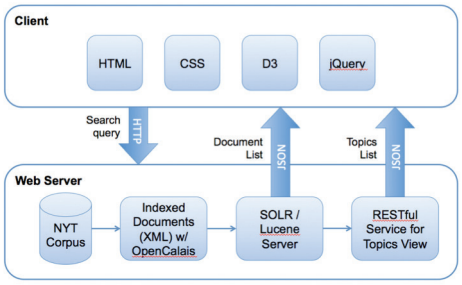

Fig 6. Footprints system architecture.

Fig. 6 shows the architecture of the system. The document set we used was New York Times Annotated Corpus [34], which included all articles published from 1987 to 2007. Our implementation used the web services from the open source topics-extraction engine called OpenCalais [35], which uses natural

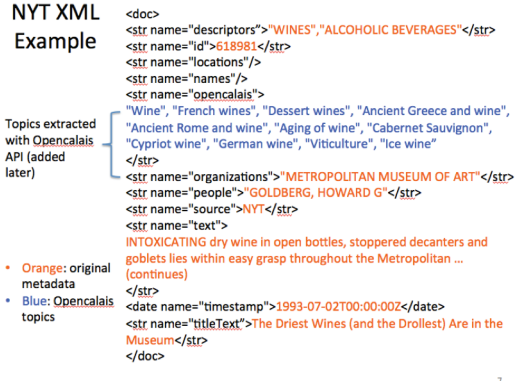

Fig 7. Sample of the NYT XML with opencalais tag appended.

language processing, machine learning, and other methods to create rich semantic metadata for a text corpus. OpenCalais extracts a set of topics contained in each article as well as the metadata category associated with each topic, such as people, places, organizations, etc.

A Java program was used to clean up the topic names by removing duplicates, trimming spaces, capitalizing topic names, and merging plural and singular versions of a topic. The XML for each document in the New York Times corpus was edited to include the "opencalais" tag, which included a comma-separated list of extracted topic names, shown in Fig. 7.

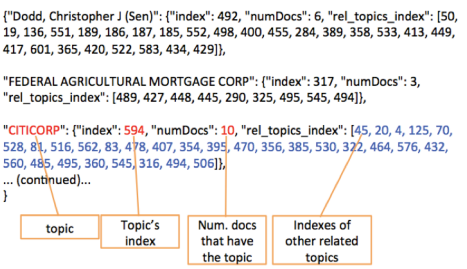

Fig 8. JSON file with metadata of each document from NYT XML.

Next, a JSON file was created that included an entry for each topic extracted from that document (Fig. 8). We assigned index ID numbers to each topic, which were included in the JSON file for each topic entry, along with a rank-ordered list of IDs for related topics, and the number of other documents in the system that also contain that topic. The list of related topics was created by pre-calculating a relatedness score between any two topics based on the percentage of documents tagged with both topics relative to the documents tagged with either one of them, as shown here:

Relatedness Score(topic a, topic b)
= Count(Intersection(A,B)) / Count(Union(A,B))

A = doc IDs of topic a
B = doc IDs of topic b

The Footprints web browser client communicates with SOLR running on our server to obtain a list of documents rank-ordered by relevance to the current search terms. It also makes an AJAX call to a RESTful service on the server to obtain a subset of the topics to be displayed in the Topics View. The algorithm for this service is discussed in section 4.1. Our server application is built using the open-source search platform Apache SOLR [33], which indexes the documents and provides access via an API for full-text search of those documents.

4.1 Topics View Algorithm

To generate the Topics View from a search term, we used the following algorithm. The system first estimates how many documents are in the corpus and how many are labelled with the topic. Then it determines how many documents were retrieved by the query and how many of those are labelled with the topic. The algorithm then uses the binomial t-test to evaluate the hypothesis that the topic is more often used in the retrieved documents compared with the overall corpus. The topic's relevance score is the Z statistic of the binomial t-test. The topics are then sorted by decreasing value of the Z score and returned in that order. The topics with negative Z scores (i.e. those that are less frequent in the target set compared to the overall corpus) are filtered out.

The use of the Z statistic to select a subset of the topics provides two useful properties:

- It takes into account the fact that topic frequency estimates are less reliable on smaller document sets, and balances the reliability of estimates with the magnitude of frequency increase.
- It gives more weight to fine-grained but rare topics relative to coarse-grained but common ones, thereby surfacing more non-obvious but potentially productive avenues of exploration.

With this approach, common topics such as "politics" that are in a great many documents would need to be exceptionally prevalent in a search result to surface as highly relevant. Extremely rare topics such as "VAST authors" would not appear simply because one of the small handful of documents happened to show up (although it would if most of them did). Instead, the algorithm prioritizes moderately unusual topics that are heavily over-represented in the results set relative to the corpus.

When the user selects a topic, the list of related topics in the JSON file is queried to get IDs above some threshold that are considered related and those topics are highlighted in blue in the Topics View. When the user adds additional terms to the search, the system constructs the Topics View as follows. First, it combines the topics with an AND operator. If the size of the resulting set is above a cut-off (set at 1,000) the algorithm terminates. If not, it progressively removes one or two query terms and runs an AND query with the remaining terms. Terms at the end of the query are removed first, under the assumption that the user had added them in order of relevance. The results from any single query are capped at 5,000 documents, both to speed up the queries and to reduce bias towards terms that appear in many documents.

This algorithm is summarized in the following pseudo-code:

```
docs = {}
For N = 0, 1, 2:
{
    For all possible ways to remove N terms from the list,
    starting with terms near the end:
    {
        - Remove N terms from the list
        - Form an AND of the remaining terms
        - Retrieve at most 5000 documents and add them to docs
    }
    If the resulting set has more than 1000 documents:
        break
}
Return docs
```

For example, if the original query included the terms 'term1' and 'term2' and the break statement was never reached, then running this algorithm is equivalent to a SOLR query of the form (term1 AND term2) OR (term1) OR (term2).

We considered this simple approach as a first attempt, just enough to get the prototype working so we could learn from it in use. We recognize that the algorithm could be improved, for example by more heavily weighting documents the user has marked as Useful, among other refinements.

5 Evaluation

Footprints' design did not emerge all at once. We designed it using a highly iterative method in which we incorporated user feedback throughout the process, both before we wrote any code and then again during implementation. During the design phase, we considered dozens of visualization approaches, rejecting many and refining others until we settled on one design. We then conducted extensive design testing using paper prototypes, which led to further refinement – including removing some of the more innovative and exploratory features that were not well received – and resulted in the design described here. Toward the end of the implementation phase, we conducted another round of evaluation to gather feedback on the working system. The following sections summarize the two phases of testing and the key lessons learned.

5.1 Design Testing

In the early design phase we took feedback from our intelligence clients (former analysts and technologists who support them), who gave us their input and passed on comments from working analysts who saw drafts along the way. Once we had a complete design, we tested it with 10 technology researchers in our own company, each time making adjustments before testing again. We then took our "pre-final" design to our client and tested it as follows.

5.1.1 Procedure

We used paper prototype testing [22] to create an extended usage scenario that involved learning about the causes of the subprime mortgage crisis, showcasing all of Footprints' features. We generated detailed design mockups that showed how the interface

appeared at each of 36 steps in the scenario. For example, one mockup showed the design with a topic box selected, the next mockup showed that topic being dragged into a search box, and the next one showed how the interface changed after it was dropped in the box, including a modified Topics View and a different number of items in the Document List.

The analysts participated individually in a two-phase process, testing first the *usability* of the design and then the *usefulness* of the features. During the usability phase, we went through the scenario one step at a time, telling the analyst what they had just "done" and asking them to interpret the changes in the interface. They were asked to think aloud about their expectations and difficulties in interpreting the interface at each step. For example, they might say, *"These boxes seem to be categories or topics of some sort, and I think the size indicates how many documents there are on that topic. The more relevant ones seem to be near the top."* If they had major misconceptions we corrected them, but mostly we let their understanding evolve unassisted. Our aim here was to see whether the design was easy to understand and learn how to improve it. In the second phase, we pointed out each of the key visualizations and features and asked the analysts to rate their usefulness on a 7-point Likert scale and explain why. The goal of this phase was to help us prioritize the features.

We tested the design with 8 intelligence analysts (A1-A8) over the course of two days, with each session lasting 90 minutes. Two participants had attended the initial requirements-gathering workshop. The sessions were video recorded. After completing the tests, we systematically analysed the videos, noting areas of confusion and transcribing their comments.

5.1.2 Results

The outcome of the testing was a long list of detailed design modifications plus a rank ordered list of feature usefulness ratings, shown in Table 1. We used the detailed feedback to modify the design to the one we implemented. Since many of those changes require an understanding of the prior design, we instead present here a high-level summary of the analysts' reactions to the features.

Rank	Feature	Avg Usefulness Rating
1	Topics-based search	6.9
2	Topics Filter	6.8
3	Coverage Histograms	6.5
4	Document Filtering	6.4
5	Document Triage	6.1
6	Topics Graph w/ Magnet Model	6.0
7	2D Search	5.8
8	Tetris Map	5.2
9	Inner Document Thumbnail	3.1

Table 1. Analysts' ratings of the usefulness of Footprints features. Scale ranged from 1-7 with 7 "extremely useful."

The analysts were extremely enthusiastic about the Topics View, specifically the idea of seeing an overview of the topics covered in the document set and creating searches based on those topics, rating it 6.9 out of 7. They saw this as a huge improvement over their current search tools, which did not contain visualizations. The following were representative of their reaction:

> *"I love that this is idea-based. It's a discovery tool for you to figure out what are similar concepts you should be looking at, or how they overlap to try to get at what you're doing." (A3)*

> *"I like the idea of it being a space that I could move around in. I think it would make me adjust my search terms. I would be like, 'Oo, that's an interesting one' and I'd drag that up to the top and have it readjust. And maybe I would be moving search terms in and out as I got closer to the area of the space that I'm really interested in, that maybe my original search terms weren't so good at reaching." (A6)*

The analysts also rated highly (6.8) the idea of filtering the Topics View to show just those topics identified as people, organizations, places, and/or things, since their briefs are frequently focused on people and their role in events and organizations. One suggestion was to provide additional filters beyond the four provided.

They also responded very well to the Coverage Histograms (rating it 6.5), both for helping them offset inherent biases in their search patterns, and for enabling them to filter based on attributes. A typical comment was:

> *"That's really useful because it's sort of a check on making sure that you're not reading just one source. I mean I couldn't even tell you right now which mainstream newspaper I read the most, just because I've never seen a visual display of it. This would be very cool. Like, am I biased? You know, 'cause we all have those internal biases that we don't often get to check." (A4)*

When probed for other attributes to display, most said Date and Source were sufficient. A few suggested that language of text or presence of media (images, video, audio) would also be useful for certain datasets.

Finally, the analysts said the Document Triage functions were necessary (rating it 6.1). Since no such triage feature is provided in their current system everyone uses a different method, including emailing documents to themselves, printing them out, and putting them in folders. One analyst said, *"If you don't include something like that, I'd somehow create a workaround, like I'd email them to myself, which is obviously not ideal."* (A4)

The remaining four features were not as well received. Ranked in the middle was a mechanism we had designed for a two-dimensional (2D) Search, which was meant to allow analysts compare concepts, see Fig. 9. A second vertical search box was added to the left of the topics view and analysts could drag a second set of topics into it to compare them to those along the top. The magnet model would pull topics up and to the left to the extent they were related to each concept, revealing the topics most relevant to both in the top left quadrant. In the evaluation, the analysts found the idea intriguing but said they needed to see it working to assess its value, so they rated it in the middle. Since paper prototyping is not as effective at testing novel features that depend on seeing changes in the data or content [22], we were not surprised by this response. Unfortunately, we ran out of time to

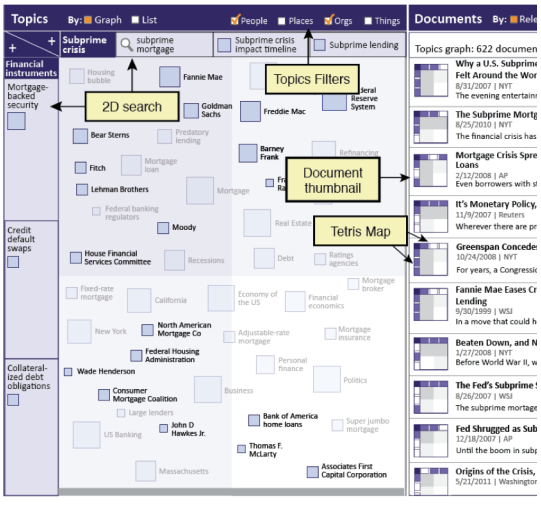

Fig 9. Original design with a 2D search in the Topics View and a thumbnail visualization next to each document.

implement it in the version we developed, so we hope to test it in a future version.

The analysts rated two more ideas fairly low. We had designed a Document Thumbnail visualization that appeared at the left of each document header showing how it was related to topics in the Topics View, shown in Fig. 9. The Tetris Map around the border of the thumbnail showed whether the document was related to each of the search topics, and the Inner Document Thumbnail indicated the degree to which it was related to the topics in each quadrant of the Topics View. Most of the analysts had trouble interpreting the thumbnail, especially the inner area, and felt it wouldn't add much value. Based on this result, we removed the thumbnail visualization from our design.

5.2 Functional Prototype Evaluation

After revising the design based on the testing, we began implementation. When the system was largely but not completely finished, we conducted another evaluation with our client to identify final adjustments and to gather input for future work.

5.2.1 Procedure

Again we showed Footprints to 8 intelligence analysts individually for 90 minutes each and video recorded the sessions. Two of the analysts had participated in the first round of testing (A2 and A7) and the other 6 were new to it (A9-A14). Again the sessions had two phases. In the first phase, we asked them to use the system in a prescribed sequence that introduced them to each of the features (and avoided certain pitfalls in the not-quite-finished prototype). In the second phase, we invited them to use Footprints to explore a topic on their own, thinking aloud as they used the system. Afterward, we interviewed them about their reactions and asked them fill out a short questionnaire.

5.2.2 Results

The analysts uniformly agreed that Footprints would be a big help to them in their research by helping them both discover topics and avoid missing things. When asked whether Footprints would *"make it easy to get a high-level overview of a topic of interest,"* they rated it an average of 6.4 on a scale of 1-7, with 7 meaning strongly agree. On the question of whether Footprints would *"help me notice topics that I might not otherwise uncover through my current search methods,"* they rated it 6.3. They gave the same score (6.3) on the question of whether the visualizations in general were *"very useful."*

These ratings were supported by their reactions while using Footprints and in the final interview. The following two quotes sum up the overall response eloquently:

> *"It's a great format. I like the combination of visualizations along with the filtering operations… It would certainly help me make connections that I didn't necessarily understand or think were intuitive. If I were looking up a certain topic – we tend to get very focused and very narrow very quick – and this would allow me to have a reference back so I can see where it fits into a larger context." (A11)*

> *"This would absolutely be useful. Taking some of these big thematic issues and trying to get an understanding of them quickly and also making sure that you thoroughly covered it is a very standard part of our job responsibilities. This seems like it's designed to support that effort." (A7).*

Once again, they were immediately enthusiastic about the Topics View and its ability to help them think of topics to search for rather than having to generate keywords, a major source of anxiety for them. The value of the histograms, on the other hand, emerged as they used the system. One analyst (A13) commented as she was using the system, "*With [my current search tool], I would not have known that in 1991 there was all this reporting – that's something that's really hard to see. So this visual part is very important."* Another analyst (A12) initially wanted to narrow the date histogram to show only the years where he thought there was activity, but as he searched he changed his mind. *"Actually, I can see that there is a benefit of having all those years there because – I guess that's the whole thing about this – if you think you're interested only in this part but all of a sudden you see that, 'oh my god, there were twice as many things back then, maybe I should go back and see what that was.'"*

This response encouraged us that Footprints was meeting its goal of supporting discoverage. Of course, there were also areas where the analysts saw potential for improvement.

5.2.2.1 Layout of the Topics View

As explained, the topic boxes in the Topics View are laid out according to the magnet model, where the topics most closely related to the search terms are pulled up toward the top. We had hoped to use the horizontal dimension for 2D searches but ran out of time. Not surprisingly, while the analysts found the magnet model easy to interpret, they expected the horizontal dimension to have meaning as well. (*"My eyes are wanting to find a pattern – is there one?"* (A7)) One good suggestion was to place newly emerging topics toward the left and declining topics to the right, which would address the requirement to help the analysts spot emerging and fading trends.

We were surprised to find that several of the analysts were bothered by the scattered layout of the topics. (*"I'm struggling a little because it feels a little scattered, without rows and columns it's hard to scan it."* (A14)) They wanted them to be displayed in a regular pattern such as a grid so that they could more easily (1) scan them, (2) return to topics of interest, and (3) refer to their locations when speaking to others. It is common for graph-like visualizations to lay out entities in a scattered pattern, so this feedback may be generally applicable to any graph-based tool where users need to scan the nodes systematically.

5.2.2.2 Filtering in both directions

Our design assumed that analysts would generate searches based on topics and then use the Date and Source histograms to filter the results. Therefore, when the user selects a bar in the Date or Source histogram, the Document List updates but the Topics View does not change so that users can keep track of other topics they might want to add to the search. The analysts clearly indicated that they expected the topics view to update when they selected a histogram bar, not just the other way around. As A13 put it, *"every single time I click around, these should be updating,"* referring to the topics and the histograms.

5.2.2.3 Getting back

The analysts also expressed concern about getting back to prior states. People mentioned wanting to get back to prior layouts of the Topics View, or prior document lists, or specific documents they had viewed earlier. Users could easily get back to documents marked Useful, but only if they had thought to mark them. We also planned to include a menu to let users return to any prior search, but we had not yet implemented it. Even so, the feedback indicated that we need to provide additional mechanisms to reconstruct prior states. Other research has focused on provenance issues involved in the search process [4, 29], but it had not emerged as a requirement in the workshop so we did not prioritize it in our design. Even at this stage, these analysts did not ask for the ability to reconstruct an entire search process; they simply wanted to return to specific prior states, generally to resume searching from that point or to re-find certain documents or topics.

Interestingly, this feedback is in tension with the analysts' previous desire for the system to update every time they make a new selection. The more dynamic the system, the harder it is for people to find things they noted in an earlier state. Our bias had been to keep the position of topics more stable so they would be easier to track, but that approach made it seem unresponsive. A

solution probably lies in providing breadcrumbs or other types of history mechanisms that allow the user to easily reconstruct prior states without having to remember to mark them. The challenge is in keeping those mechanisms simple and uncluttered with every possible prior state.

5.2.2.4 More refined Useful tag

While the Useful tag was seen as essential, it wasn't sufficient. The analysts are frequently tracking multiple topics and they wanted to categorize the documents marked Useful. Our intention was to use document tags for this purpose and we showed the analysts how this might work, since it was only partially implemented (see Fig 1). But we could see that our design did not effectively integrate the tagging feature with the simpler Useful feature. Perhaps a list of tags could be accessible from the Useful star to the left of the document title.

The analysts also wanted the opposite of a Useful feature, namely the ability to "Ignore" both topics and documents. Since they are ever-concerned about missing things, they did not want to remove them entirely; they simply wanted to move topics off to the side or hide documents in the results list. And just as the Topics View algorithm should weight documents marked Useful more heavily, it should also lower the weight of Ignored documents.

6 Conclusion

We have introduced a tool call Footprints that is designed to support both discovery and coverage, or discoverage. It provides cues to help analysts discover where they should be looking in several ways. The topics view shows only the most relevant topics extracted from the document set rather than the full space of topics, and it surfaces more fine-grained topics that tend to be helpful for uncovering unanticipated topics. It makes it easy to identify certain types of topics, such as people, places, and organizations. The topics view also highlights topics highly related to selected documents, topics, or both, suggesting other promising topics to explore. The date and source histograms also give hints about where to look by showing when a topic was discussed by whom over time.

Footprints supports coverage tracking by showing the proportion of documents the user has read by date, source, and topic. The coverage histograms make it easy to filter the documents in complex ways to fill in any gaps in coverage. The persistence of the coverage markers across searches in a session lets the user see when her searches are no longer uncovering many new documents so she knows when she can stop searching. They also help her quickly identify documents similar to ones she previously found useful. These coverage indicators could also be extended to a community by showing people which documents were read and found useful by others. To be clear, the coverage histograms cannot point analysts to specific key documents on a subject, but they can help them notice and correct any biases in their coverage along certain dimensions – or at least be aware of those biases.

Two rounds of evaluation with analysts indicated that Footprints succeeded in its two main objectives of helping them discover relevant information even when they are not sure what to look for, and knowing how well they have covered the related material so they know when they can stop searching. It was striking to us how much the analysts responded to the simple idea of visualizing the topics underlying a document collection and using them to generate search queries. Within the research community the idea of visualizing the topic structure is commonplace, but it had not been deployed to these analysts. Similarly, the analysts were delighted to see the distribution of documents in an interactive histogram, and more so, impressed by the power of seeing their coverage overlaid on it. Still, testing indicated that Footprints could be improved by using the horizontal dimension of the Topics View to show the "freshness" of topics as they emerge and decline in the news, by making it easier to return to prior states, by enhancing the tagging feature, and by allowing users to filter the Topics View by date range and sources.

We designed Footprints using a highly iterative, user-centered approach in which we systematically tested the design with users during both the design and the implementation phases and carefully analyzed the feedback. The design underwent dramatic transformations based on this process, and in some cases led us to discard more novel visualizations that were not well received. Although it was often disappointing to let go of those ideas, we appreciate that realistic user feedback helped us stay focused on providing the most effective visualizations and features for an applied usage setting. As a result of the analysts' responses, our client is currently working to integrate aspects of Footprints into the analysts' suite of tools, and we hope to learn about its impact as it is deployed and incorporated into daily use.

Acknowledgments

We thank Ignacio Solis and Oliver Brdiczka for their valuable contributions to this project. We are especially grateful to our government clients for their guidance and assistance and to the researchers and analysts who participated in the evaluations.

References

[1] Chi, E.H., Pirolli, P., Chen, K., & Pitkow, J. (2001). Using information scent to model user information needs and actions and the Web. *Proc. of Computer-Human Interaction*. ACM, 490-497.

[2] Chuang, J., Ramage, D., Manning, C. D., and Heer, J. (2012). Interpretation and Trust: Designing Model-Driven Visualizations for Text Analysis. *Proc. of Computer-Human Interaction*. ACM, 443-452.

[3] Collins, C., Viegas, F.B. and Wattenberg, M. (2009). Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora. *Symposium on Visual Analytics Science and Technology*, IEEE, 91-98.

[4] Gotz, D. & Zhou, M. (2009). Characterizing Users' Visual Analytic Activity for Insight Provenance. *Information Visualization* 8(1): 42-55.

[5] Gretarsson, B., O'Donovan, J., Bostandjiev, S., Llerer, T.H., Asuncion, A., Newman, D., and Smyth, P. (2012) TopicNets: Visual Analysis of Large Text Corpora with Topic Modeling, *Transactions on Intelligent Systems and Technology,* ACM, 1-26.

[6] Havre S., Hetzler, E., Perrine, K., Jurrus, E., and Miller, N. (2001) Interactive Visualization of Multiple Query Results. *Proc. of Information Visualization,* IEEE, 105-112.

[7] Hetzler E. and Turner A, (2004). Analysis Experiences Using Information Visualization, *Computer Graphics and Applications,* IEEE, 22-26.

[8] Heuer, R. (1999). Psychology of Intelligence Analysis. United States Government Printing.

[9] Johnston, R. (2005). Analytic Culture in the US Intelligence Community: An Ethnographic Study. Analysis. Central Intelligence Agency, Washington, DC.

[10] Kang, H., Plaisant, C., Lee, B., and Benderson, B., (2007). NetLens: Iterative Exploration of Content-Actor Network Data, *Proc. of Information Visualization,* IEEE, 18-31.

[11] Kim, K., Ko, S., Elmqvist, N., and Ebert, D.S. (2011) WordBridge: Using Composite Tag Clouds in Node-Link Diagrams for Visualizing Content and Relations in Text Corpora. *Proc. of Hawaii International Conference on System Sciences,* IEEE, 1-8.

[12] Kwon, B. C., Javed, W., Ghani, S., Elmqvist, N., Yi, J. S., & Ebert, D. (2012). Evaluating the Role of Time in Investigative Analysis of Document Collections. *Transactions on Visualization and Computer Graphics,* IEEE, 18(11):1992-2004.

[13] Lamping, J., Rao, R., and Pirolli, P. (1995) A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies. *Proc. of Computer-Human Interaction,* ACM, pp. 401-408.

[14] Lee, B., Czerwinski, M., Robertson, G., and Bederson, B.B. (2004). Understanding Eight Years of InfoVis Conferences Using PaperLens. *Proc. of Symposium on Information Visualization*, IEEE, 216.3.

[15] Marrin, S. (2003). CIA's Kent School: Improving Training for New Analysts. *International Journal of Intelligence and Counter Intelligence,* 16, 609–637.

[16] Marrin, S. (2002). Homeland Security and the Analysis of Foreign Intelligence, Markle Foundation Task Force on National Security in the Information Age.

[17] Patterson, E. S., Roth, E. M., & Woods, D. D. (2001). Predicting Vulnerabilities in Computer-Supported Inferential Analysis Under Data Overload. *Cognition Technology and Work*, 3, 224-237

[18] Patterson, E. S., Woods, D. D., Tinapple, D., Roth, E. M., Finley, J., & Kuperman, G. G. (2001). Aiding the Intelligence Analyst in Situations of Data Overload: From Problem Definition to Design Concept Exploration. *Institute for Ergonomics/Cognitive Systems Engineering Laboratory Report,* ERGO-CSEL, 1-105.

[19] Pirolli, P. and Card, S. (2005). The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis, *Proc. of International Conference on Intelligence Analysis,* McLean, VA, Vol 6.

[20] Rennison, E. (1994). Galaxy of News: An Approach to Visualizing and Understanding Expansive News Landscapes. *Proc. of User Interface Software and Technology,* ACM, 3-12.

[21] Rzeszotarski, J. & Kittur, A. (2014) Kinetica: Naturalistic Multi-touch Data Visualization. *Proc. of Computer Human Interaction,* ACM: 897-906.

[22] Snyder, C. (2003). *Paper Prototyping*, San Francisco: Morgan Kaufmann.

[23] Spence, R. (2007). Information Visualization: Design for Interaction, Prentice Hall.

[24] Spoerri, A, (2004). RankSpiral: Toward Enhancing Search Results Visualizations, *Proc. of Information Visualization*, IEEE, 215-216.

[25] Stasko, J., Görg, C., & Spence, R. (2008). Jigsaw: Supporting Investigative Analysis Through Interactive Visualization. *Information Visualization,* 7(2), Nature Publishing Group, 118-132.

[26] Thomas, J. J., & Cook, K. A. (2005). Illuminating the Path: The Research and Development Agenda for Visual Analytics, *Annals of Physics* (Vol. 54). IEEE.

[27] Van Ham, F., Wattenberg, M. and Viegas, F.B., (2009). Mapping Text with Phrase Nets, *Transactions on Visualization and Computer Graphics,* IEEE, 1169-1176.

[28] Viegas, F.B., Wattenberg, M., van Ham, F., Kriss, J. and McKeon, M. (2007). ManyEyes: a Site for Visualization at Internet Scale. *Transactions on Visualization and Computer Graphics*, 13, 6, IEEE, 1121-1128.

[29] Wen, Z., Zhou, M. & Aggarwal, V. (20007). Context-Aware, Adaptive Information Retrieval for Investigative Tasks, *Proc. Of IUI*, ACM, 122-131.

[30] Willet, W., Heer, J., & Agrawala, M. (2007). Scented Widgets: Improving Navigation Cues with Embedded Visualizations. *Transactions on Visualization and Computer Graphics*, IEEE, 13(6):1129-1136.

[31] Yi, J. S., Melton, R., Stasko, J., & Jacko, J. A. (2005). Dust & Magnet: multivariate information visualization using a magnet metaphor, *Information Visualization*, 4(4), 239-256.

[32] http://in-spire.pnl.gov

[33] https://lucene.apache.org/solr

[34] http://catalog.ldc.upenn.edu/LDC2008T19

[35] http://www.opencalais.org