From Question Answering to Visual Exploration

David McColgin
Pacific Northwest National Laboratory
902 Battelle Blvd
Richland, WA. 99354
509-375-2148

Michelle Gregory
Pacific Northwest National Laboratory
902 Battelle Blvd
Richland, WA. 99354
509-375-2824

Elizabeth Hetzler
Pacific Northwest National Laboratory
902 Battelle Blvd
Richland, WA. 99354
509-375-6698

Alan Turner
Pacific Northwest National Laboratory
902 Battelle Blvd
Richland, WA. 99354
509-375-6670

{david.mccolgin;michelle.gregory; beth.hetzler;alan.turner}@pnl.gov

ABSTRACT

Success in Question Answering has been traditionally measured by precision and recall, which are good metrics for identifying specific best answer(s) that might be obtained by a lookup type of search. These metrics do not address the many information gathering techniques in exploratory interactions. In this paper, we present an integrated Question Answering environment that combines a visual analytics tool with state-of-the-art query expansion, and complements the cognitive processes associated with an information analyst's work flow. In our system, questions result in a comprehensive answer space that allows users to explore the variety within the answers and spot related information in the rest of the data. The exploratory nature of the dialog between the user and this system requires tailored evaluation methods that better address the evolving user goals and counter cognitive biases inherent to exploratory search tasks.

Categories and Subject Descriptors

I.3.6 [**Computer Graphics**]: Methodology and Techniques – Interaction techniques, I.6.9 [**Visualization**] – Information Visualization, Visualization Techniques and Methodologies, I.7.5 [**Document Capture**] - Document analysis, H.5.2 [**User Interfaces**] - Evaluation/methodology, H.3.3 [**Information Search and Retrieval**] - Search process

General Terms

Design, Human Factors.

Keywords

Information Visualization, User interaction design, exploratory search, evaluation.

1. INTRODUCTION

Marchionini defines three types of search: lookup, search to learn, and investigative [6]. While returning factoid answers satisfies many search needs, the information needs of an information analyst require an investigative approach. In this paper we present an integrated Question Answering (QA) system that combines state-of-the-art query expansion [2] with a document visualization tool, IN-SPIRE [3]. In this system, users query a document space with a natural language question that is expanded and optionally edited by the user. Queries result in the identification of relevant passages and the selection of matching documents within the context of the whole document set. This approach leads to a sophisticated dialog in which the user can explore the QA results and maximize understanding of the data before reading individual documents and without relying solely on retrieved passages. The advantage of this analysis environment lies not in the power of any one visualization or tool, but in the process supported by using them in concert. With improved understanding of the answer space, users can better form new questions, detect answer patterns, or select the most interesting documents to read in detail. The system we present here supports analysts' goals by helping to identify the presence of conflicting data, data from other sources, answer patterns (e.g. geographical or temporal), and even information on other topics not returned by the query but potentially relevant. The evolving information needs of the analysts require system evaluation metrics that go beyond precision. In this paper, we discuss the information needs of analysts and use a work flow scenario to present our exploratory system. We report on initial formative evaluation of the system and conclude with a discussion of formal, summative evaluation metrics.

2. THE ANALYST'S TASK

Information analysts spend much of their time foraging complex and contradictory bodies of information in support of their ultimate reasoning goals. They are seeking detailed knowledge of specific facts that can 1. support or refute candidate positions on the subject they are investigating; 2. allow them to credibly identify and bridge the gaps in their knowledge, and; 3. discover previously unknown evidence and relationships. As domain experts on the topics that they are exploring, their goal is not to simply isolate "the best" facts, but rather to explore new dimensions of the data and arrive at reasoned and supportable conclusions [9]. They perform these tasks under significant pressures and constraints including time limits, the required form of their output (e.g., a verbal briefing, a written report, the length of the report, etc.), often unfamiliar topics and great uncertainty, and information sources of variable accuracy.

An exploratory system for information analysts must maximize data understanding within the level of domain knowledge and multiple constraints on the working conditions. In addition, such a system also needs to help overcome the potential cognitive pitfalls of analytical work under pressure such as satisficing, anchoring, vividness, and oversensitivity to consistency [1,4].

QA and interactive query expansion within a visual analytics environment offer the chance to counteract such biases. Instead of querying, interpreting limited results, and querying again, the analyst is presented with a comprehensive visual answer space that can be interactively explored. Variations in the extracted answer passages, contextual information about other documents in the collection, and patterns in the answers across time, source, or theme reveal alternative explanations and unanticipated influential factors.

3. IN-SPIRE

IN-SPIRE is a visual analytics tool developed by Pacific Northwest National Laboratory to facilitate rapid understanding of large textual corpora [3]. IN-SPIRE generates mathematical signatures for each document in a set. Document signatures are clustered according to common themes to enable information exploration and visualizations. Information is presented to the user using several visual metaphors to expose different facets of the textual data. The central visual metaphor is a Galaxy view of the documents as clustered dots that allows users to intuitively interact with thousands of documents, examining them by theme (Figure 1). The Galaxy has been shown to provide value beyond traditional retrieval systems [3]. While the concept of cluster projection is not new, the current line of research is exploring its value within a larger visual QA process. Additional analytic tools allow exploration of temporal trends, thematic distribution by source or other metadata, and query relationships and overlaps.

QA functionality is being integrated within IN-SPIRE so that users can explore specific questions within the massive data collections. Query expansion is provided by Language Computer Corporation's FERRET application [2]. The interface is incorporated into IN-SPIRE within its query tool. Users can ask questions in natural language and FERRET finds answer passages as well as returning expanded queries in Boolean and Query by Example (QBE) syntax for use in IN-SPIRE. Users have direct access to the output and can edit or add terms to the query. Queries can search the whole dataset or only the currently selected documents to help refine an information need. In the next section, we demonstrate these capabilities through a sample scenario.

4. WORK FLOW SCENARIO

An analyst is given the task to determine the largest environmental threats posed by nuclear power. Given the plethora of avenues one can use to form hypotheses on this topic, simple searches with factoid answers are not adequate to explore all relevant details from the data. The first step in the work flow is exploring the document collection within which the analyst will work. The dataset can be opened in the Galaxy view (see Figure 1), which allows her to assess the size and thematic coverage of the collection.

The task could be approached in a variety of ways but she would first like an historical perspective. She asks "Which nuclear reactors had the worst safety incidents?" in IN-SPIRE's Query tool. FERRET returns documents with answer passages, giving her concise facts about her question and helping to guide her subsequent investigation (Table 1).

A malfunctioning control rod caused the shut down of Zaporizhzhya-4 on 13 April 1997. [1, 2] According to a **plant** spokeswoman, one of the 61 control rods used to moderate **nuclear** activity failed to descend into the **reactor** core within the time allowed by regulations.

Table 1. Sample answer.

Expanded queries are also returned below the original question in both Boolean and vector-based forms (Table 2). Together, the two expansions provide syntactic expansion, top-down semantic expansion based on external sources like WordNet, and bottom-up instances of related terms from the data itself.

(reactors OR "nuclear reactor" OR reactor) AND (worst OR defective OR risky) AND (incidents OR matter OR event OR incident) AND (safety OR safe OR guard) AND (nuclear OR atomic)

Table 2. Sample Boolean expansion.

Analytically, the query expansions help to start a dialog that provides additional insight and context. The analyst reviews the expansions and decides to use the Boolean query. She has the chance to remove undesired terms, change the Boolean logic, and add concepts of her own before executing.

4.1 Galaxy View

Her results appear ordered by relevance in the Document Viewer where she can access the title and full text. Results are also displayed in the Galaxy in the context of the entire collection. The clusters and labels help her gain valuable insight into the content of the query results without having to read each document individually or rely solely on the documents that match her query. The thematic view helps her to identify and eliminate irrelevant results, refine her information need, or find a new facet of the answers worth exploring. Nearby labels describe the related topics and nearby documents contain related material that the query results alone would not have provided. In this case, the analyst decides to visualize the result documents alone (Figure 1).

When the Galaxy is recalculated to show only the results, she begins to investigate the clusters. Recognizing the name Chernobyl she first scans documents in clusters with that label and finds information about several specific incidents, their effects, and the international response. Investigating the clusters to the right labeled "integrum, mayak, ctr" and "launchers, ss, india", she gains insight into an unexpected risk; these documents contain information on accidents and radiation leaks from military vehicles such as ships and submarines, the threat posed to marine environments, and the remediation efforts and methods in use. While they do not directly answer her question, these documents provide complementary information about environmental safety. When she examines the documents clustered at the bottom of the visualization, another theme emerges. In contrast to the clusters above, these documents are primarily about new reactors and development programs with much discussion on new safety technologies and protocols that could mitigate the risks and effects of future incidents.

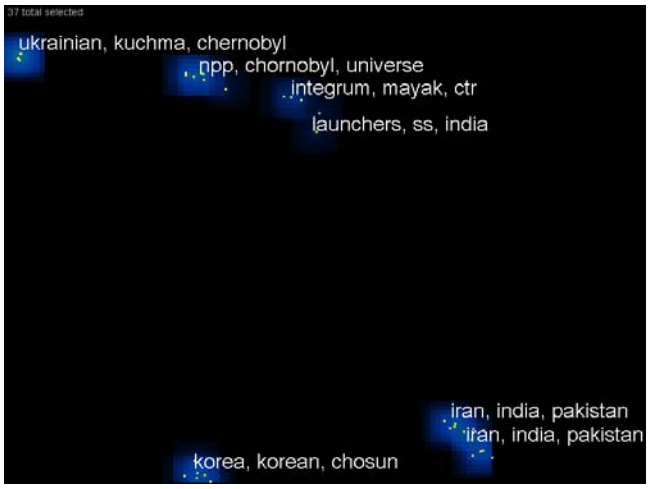

Figure 1: Galaxy view recalculated to include only the 37 query results.

4.2 Correlation Tool

The analyst can group any set of selected documents, whether they are selected manually or by virtue of matching a query. In this example, she has made many groups based on the answers, Boolean query, selections from the other tools, and independently determined groups such as countries. The Correlation tool allows her to explore the overlap between the groups she has created. Figure 2 shows her query results (y-axis) distributed over countries (x-axis). As with all of the tools, Correlation is linked to the visualizations, so that clicking a column here results in a selection in the other tools.

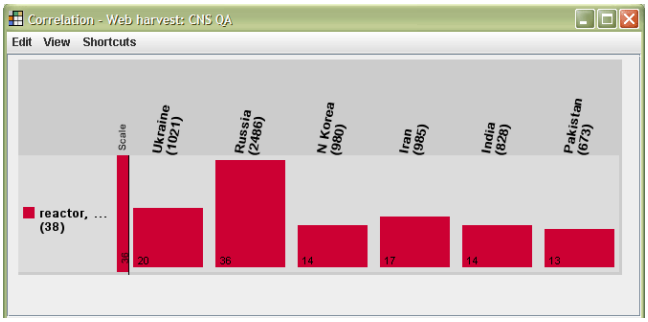

Figure 2: Distribution of expanded query results by country.

4.3 Additional Tools

There are a variety of other interactive tools and visualizations that can help analysts investigate and gain insight in IN-SPIRE. Affect, trends and salient events in time, and other data attributes can be measured, portrayed, and used interactively. Hypothesis tracking and outlier evolution are also explicitly supported.

4.4 Review

By now, the analyst has an overall sense of the data from the visualization, which helped her to formulate an initial question. She was presented with extracted answers that helped her refine the system's query expansions to her interests. Portraying query results in context helped her find useful non-hits that could be important. Seeing variability within her results allowed her to find unanticipated relevant information: an unexpected type of risk and information on modern safety improvements. She was able to see the country with the most content about safety incidents using the Correlation tool, and see differences over time. Certainly, the process helps the analyst find facts and discover evidence and relationships. But it also exposes facets of the answer space that influence her subsequent interactions and helps to further define her information need. The analyst may continue exploring by refining her question, starting a new line of inquiry based on her acquired knowledge, or performing deeper analysis with the other tools.

5. EVALUATION

The challenge of evaluating exploratory search systems shares many of the challenges of evaluating visual analytics systems in general. Certainly, *usability* is one part of the solution, including quantitative measures, such as time and errors, and qualitative measures, such as user satisfaction. In the case of our tool, formative usability evaluation with analysts has helped reinforce our main direction while suggesting specific improvements, such as additional kinds of user interaction.

Several approaches to evaluating *utility* also provide merit, although exploratory utility is harder to assess given the lack of solid "correct answers." NIST used quality of users' written analysis reports as one metric to evaluate the system used in creating the report [5]. Contests such as those run by the InfoVis Symposium judge systems based on the ability of the tool to interactively reveal insights into the data. [8] We have also found that having a tool developer or designer work together with a user to carry out an analysis task can be an excellent way not only to assess the potential utility of the system, but also to sharpen the perception of user needs.

We propose that a good exploratory system should encourage *sound usage strategies*, and are researching an approach where this goal serves as the basis of evaluation. For example, an experiment conducted by Patterson et al. identified searching behaviors that led to exclusion of key documents, correlating very well with errors in users' verbal reports [7]. Typically, users started with a broad search and then progressively narrowed it to reduce the number of hits to a reasonable level, often excluding key documents without realizing it. In addition to the obvious metric of how many key documents were found, several complementary metrics could provide insight:

1) How many search paths did the user try? It's routinely easy for users to add terms to a previous query, often to narrow the results; fewer systems make it easy to try a new tack or combine multiple strategies.
2) Of the key documents found, how many were recognized as important? This is a subtle question, aimed at assessing a system's capabilities to quickly help users assess the value of documents. Many systems provide metadata, such as year and source, or fragments of text to help with this assessment. Still, the daunting task of skimming tens or hundreds of such fragments may lead users to quickly resort to a new smaller search. What kinds of clues are needed to ensure that once a key document is located (e.g., by a search), it is actually recognized and not discarded?
3) How many of the key documents were actually considered in the user's decisions? In the midst of information overload, users may easily forget details of specific documents. Exploratory systems provide a challenge in this regard, as the information tasks often follow unexpected paths. This metric is aimed at evaluating how well a system supports retention and use of important discovered information.

Another facet of sound usage is the ability of a system to help *counter user bias*. Exploratory search systems are inherently a partnership between user and system, and ideally should utilize the strengths of each to compensate for the weaknesses. As discussed in Section 2, users carrying out investigative search are vulnerable to a number of cognitive biases. In contrast to a lookup search task for which there may be a single best answer, an investigative task involves identification and consideration of multiple alternative answers. One example bias is anchoring,

where a user's initial judgment or estimate of the answer unduly influences evaluation of subsequent evidence [4]. Metrics related to this bias might include:

1) How many alternatives did the user explore? This question tries to go beyond the simple identification of alternatives to assess whether the user spent time actually investigating more than one explanation. A system's ability to support and track multiple alternatives can make this task easier, hopefully leading to more in-depth investigations by users.
2) How much credence did the user give to counter evidence? This question aims at one of the aspects of anchoring, that users will discount evidence contrary to a chosen explanation.

While enticing, time as a metric can be misleading in this context. A system that helps users more quickly come to a conclusion might also contribute to anchoring or satisficing rather than helping to counter them.

Exploratory systems can provide great value to users in many fields. While recognizing the value of usability and utility measures, we propose that metrics be developed based on sound usage strategies and the combination of user/system capabilities to counter weaknesses in each.

6. ACKNOWLEDGMENTS

IN-SPIRE developers Jon McCall and TJ Hoeft were key contributors to the technical implementation. Our thanks also to the Language Computer Corporation and especially Sanda Harabagiu, John Lehman, and Patrick Wang for their collaboration on integrating FERRET.

7. REFERENCES

[1] Connaway, L.S., Prabha, C., & Dervin B. An overview of the IMLS Project "Sense-making the information confluence: The whys and hows of college and university user satisficing of information needs". Presented at Library of Congress Forum, American Library Association Midwinter Conference, Boston, MA, Jan 16, 2005.

[2] Harabagiu, S., Hickl, A., Lehmann, J., & Moldovan, D.. Experiments with interactive question-answering. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05) (Ann Arbor, June 25-30, 2005). Association for Computational Linguistics, New Brunswick, NJ, 2005, 205-215.

[3] Hetzler, E. and Turner A. Analysis experiences using information visualization. IEEE Computer Graphics and Applications, 24:5, 2004.

[4] Heuer, R.A. *Psychology of Intelligence Analysis*. Government Printing Office, Pittsburgh, PA, 1999.

[5] Kelly, D., Kantor, P.B., Morse, E.L., Scholtz, J., Sun, Y. User-centered evaluation of interactive question answering systems. In Proceedings of the Interactive Question Answering Workshop at HLT-NAACL (New York City, NY, June 8-9, 2006). Association for Computational Linguistics, Stroudsburg, PA, 2006 49–56.

[6] Marchionini, G. Exploratory search: from finding to understanding. Communications of the ACM. 49:4 2006, 41-46.

[7] Patterson, E.S., Woods, D.D., Tinapple, D., Roth, E.M., Finley, J.M., and Kuperman, G.G. Aiding the Intelligence Analyst in Situations of Data Overload: From Problem Definition to Design Concept Exploration. Technical Report ERGO-CSEL 01-TR-01, Inst. for Ergonomics/Cognitive Systems Engineering Lab, Columbus, OH, 2001.

[8] Plaisant, C. The challenge of information visualization evaluation. In Proceedings of the Working Conference on Advanced Visual Interfaces. ACM Press, New York. 2004, 109-116.

[9] Thomas, J. and Cook, K.A. (Eds.). *Illuminating the Path--Research and Development Agenda for Visual Analytics*. IEEE Press, Los Alamitos, CA, 2005.