Capturing, Visualizing, and Sharing the Process of Data Analysis

Ritesh Agrawal[1], William Pike[2]

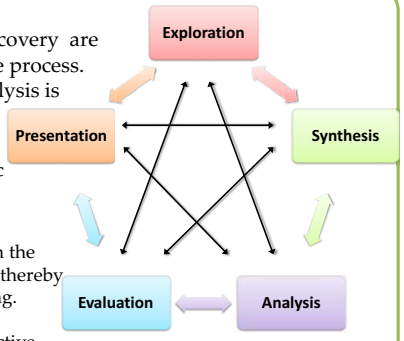
[1]Department Of Geography, North-East Visualization & Analytics Center, The Pennsylvania State University

[2]Pacific Northwest National Laboratory

Motivation

Data analysis and knowledge discovery are complex, iterative and a subjective process. Capturing the process of data analysis is important to:

- Help an analyst keep track of human thoughts, important analytic activities, biases and uncertainty in the discovered knowledge.
- Situate the discovered knowledge in the context in which it was created and thereby facilitate effective knowledge sharing.
- Help system designers to build effective tools by identifying and automating commonly used patterns in knowledge discovery.

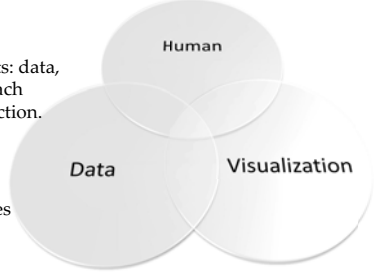


Nexus of activities comprising the scientific process (Gahegan, 2004)

Challenge

Capturing the process of data analysis is challenging for the following reasons:

- In a visual analytic environment, data analysis involves different components: data, visualization, and human thoughts; each operating at a different level of abstraction.
- Human thinking does not always prompt changes in visualization. Similarly, changes in visualization does not always prompt changes in data. Thus, transition in human thoughts, visualization, and data occur at different time intervals.



Components of a visual analytic environment

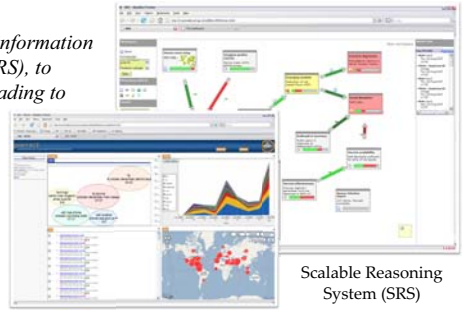
- At each level of abstraction, there are different parameters that must be stored:
  - **Data**
    - Reference to datasets that were used during analysis
  - **Visualization**
    - Visual Settings
    - Synchronization settings between views
    - Human interactions with views
  - **Human Cognition**
    - Annotations – Textual/Visual/Audio
    - Task, plan, strategy, etc.

A Prototype

The goal of this work was to extend an existing information analysis tool, the Scalable Reasoning System (SRS), to capture and represent the cognitive processes leading to knowledge discovery

1 Introduction to the Scalable Reasoning System

- Web-based visual analytic environment
- Supports end-to-end analytic activities
- Service-Oriented-Architecture (SOA)
- Organizes facts, information, and knowledge as nodes and links

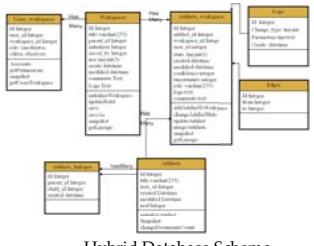


Scalable Reasoning System (SRS)

2 Capturing Analytic Activities

a. **Identify actions** supported by SRS and determine their equivalent cognitive tasks

b. Develop a **hybrid database** schema to store state of data and system over time



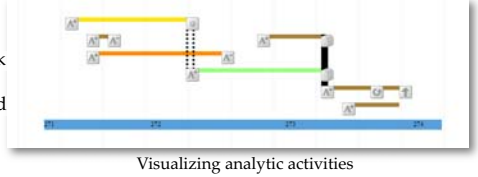
Hybrid Database Schema

	Actions	Cognitive Tasks
1.	Create an artifact	An information unit to keep track of
2.	Delete an artifact	Information that is no longer relevant or important
3.	Add documents to an artifact	Identifying documents belonging, supporting, or related to an information unit
4.	Remove documents from an artifact	Documents that seem irrelevant to the overall information contained by an artifact
5.	Merge artifacts	Group two different kinds of information
6.	Create a new child artifact	Subset of overall information that needs to be separately highlighted, saved, or analyzed
7.	Convert a node to a link	This information provides a link between two information units
8.	Convert a link to a node	This information no longer provides a linkage between two different information
9.	Add a source node to a link	This information provides support in the context of information provided by the link.
10.	Add a target node to a link	This information is derived in the context of information provided by the link.
11.	Remove source node of a link	This information no longer supports the target information in the context of information provided by the link.
12.	Remove target node of a link	This information can no longer be derived in the context of information provided by the link.

Actions supported by SRS

3 Visualizing Analytic Activities

- A timeline based visualization is used to represent analytic activities
- Each icon represents one of the analytics task from the above table. Each row in the timeline represents different tasks performed on a single information artifact
- Each line represents the duration over which an information artifact persist.
- Line color indicate state of an information artifact, such as, assumption, hypothesis, fact, resource, annotation, etc.



Visualizing analytic activities

Summary

Recent advances in visual analytics facilitate data analysis and knowledge discovery. However keeping track of important analytic activities and human thinking necessary to analyze, evaluate, and share discovered knowledge is still a challenge. This research explores these challenges and develops a prototype solution to capture, visualize, and analyze the process of data analysis.

By identifying visual interactions and actions afforded by SRS and establishing relationships between these actions and cognitive tasks, an infrastructure to capture human thinking is developed. Further, a timeline visualization is implemented to represent important decisions of an analyst and the context in which these decisions were made.

References

Gahegan, Mark. "Beyond Tools: Visual Support for the Entire Process of GIScience." Exploring Geovisualization. Ed. J. Dykes, Alan M. MacEachren and M. J. Kraak. New York:Pergamon, 2005. 75-91.

Pike, William, Richard May and Alan Turner. "Supporting Knowledge Transfer through Decomposable Reasoning Artifacts." 40th Annual Hawaii International Conference on System Sciences. 2007.

Acknowledgement

This project was funded through the Regional Visualization and Analytics Centers (RVACs) by the Department of Homeland Security, Science and Technology Directorate, Office of University Programs.

Contacts

Ritesh Agrawal (ragrawal@psu.edu)
The Pennsylvania State University

Dr. William Pike (william.pike@pnl.gov)
Pacific Northwest National Laboratory

Dr. Richard May (richard.may@pnl.gov)
Pacific Northwest National Laboratory