Analytic Provenance: Process + Interaction + Insight

Chris North
Virginia Tech
2202 Kraft Dr.
Blacksburg, VA 24060
north@cs.vt.edu

Remco Chang
Tufts University
161 College Ave.
Medford, MA 02155
remco@cs.tufts.edu

Alex Endert
Virginia Tech
2202 Kraft Dr.
Blacksburg, VA 24060
aendert@cs.vt.edu

Wenwen Dou
UNC Charlotte
9201 University City Blvd,
Charlotte, NC 28223
wdou1@uncc.edu

Richard May
Pacific Northwest National Lab
902 Battelle Boulevard
P.O. Box 999, MSIN J4-32
Richland, WA 99352
Richard.may@pnl.gov

Bill Pike
Pacific Northwest National Lab
902 Battelle Boulevard
P.O. Box 999, MSIN J4-32
Richland, WA 99352
Bill.pike@pnl.gov

Glenn A. Fink
Pacific Northwest National Lab
902 Battelle Boulevard
P.O. Box 999, MSIN J4-32
Richland, WA 99352
Glenn.fink@pnl.gov

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. One key aspect that separates visual analytics from other related fields (InfoVis, SciVis, HCI) is the focus on analytical reasoning. While the final products generated at from an analytical process are of great value, research has shown that the processes of the analysis themselves are just as important if not more so. These processes not only contain information on individual insights discovered, but also how the users arrive at these insights. This area of research that focuses on understanding a user's reasoning process through the study of their interactions with a visualization is called *Analytic Provenance*, and has demonstrated great potential in becoming a foundation of the science of visual analytics. The goal of this workshop is to provide a forum for researchers and practitioners from academia, national labs, and industry to share methods for capturing, storing, and reusing user interactions and insights. We aim to develop a research agenda for how to better study analytic provenance and utilize the results in assisting users in solving real world problems.

Keywords
Analytic provenance, user interaction, visual analytics, visualization

ACM Classification Keywords
H5.m. Information interfaces and presentation

General Terms
Human Factors

Introduction
Understanding a user's analytic reasoning process when using a visual analytics system has become an important research topic in the visual analytics community. Central to the mission of the visual analytics research agenda [1], this research aims at understanding how a user interacts with a visual interface to perform analytical tasks. With such an understanding, researchers and developers can design better interfaces that assist reasoning flow, enable knowledge sharing, and eventually support human-computer mixed initiative systems [1].

Although recent research has shown that a user's reasoning process can be retrieved through examination of a user's interaction history [2], there is little agreement on how to best capture a user's interactions, store the user history, or retrieve the user's reasoning process. Researchers in various domains have designed and implemented proprietary mechanisms that are suitable for their domains (such as automatic tutorial generation [3], scientific visualization [4], network detection [5], etc.), but it is largely unclear how the success of one system can be applied to a different system in an unrelated domain. The goal of this workshop is to bring these researchers and practitioners together to share their experiences, and discuss what steps are necessary for developing a deeper understanding of analytic provenance as both a theory and a practice.

Background
A central precept of visual analytics is that the development of human insight is aided by interaction with a visual interface, and the steps that a user takes to discover insights are often as important as the final product itself [6]. The key to the research of analytic provenance is the belief that by capturing a user's interactions with a visual interface, some aspects of the user's reasoning processes can be retrieved. In practice, we propose that the research of analytic provenance can be examined in five interrelated stages: *perceive*, *capture*, *encode*, *recover*, and *reuse*.

Perceive
In order to correlate a user's interactions with a visualization to her reasoning process, the research must begin with understanding how the data is presented to the user. As shown by Dou et al., combining the visual representation with the interaction history can disambiguate "why" a user performs certain interactions [2]. Since the user's interaction can only begin after perceiving the visualization of data, the analytic provenance research also needs to start with the understanding of how information is perceived by the user.

Capture
As the user interacts with visualization, the series of interactions can be considered as a linear sequence of actions. The most common application of this concept is the use of "undo" and "redo" buttons that are available to most computer software today [7]. However, such information is often insufficient in representing the user's reasoning process. Researchers have shown that additional semantic information is necessary to adequately represent a user's analysis

process [6]. Such semantic information can be directly annotated by the user [8], modeled based on task analysis [9], or correlated with the visualization elements [2], but identifying the most appropriate representation remains an open challenge [10].

Encode
Encoding refers to the process of describing the captured provenance in predefined formats. While many systems implicitly have their own encoding schema for capturing analytic provenance for specific tasks and domains, few generalizable schemas exist. Researchers have attempted using XML [11], declarative pattern language [5], logic-programming [12], and dynamic scripts [13], but in most cases these schemas only record the "how", but not always the "why". By using these schemas, the user can reapply interaction, but the semantic meanings behind these steps are often unclear.

Recover
Once the user's provenance has been captured and encoded, the challenge becomes making sense of the provenance. As noted by Jankun-Kelly et al., history alone is not sufficient for analyzing the analytical process with visualization tools [11]. Often, there are relationships between the results and other elements of the analysis process which are vital to understanding. While some of the relationships have been shown to be recoverable through manual inspection [2], whether the same can be done using automated techniques is still an open question.

Reuse
One important goal of the research in analytic provenance is to be able to automatically reapply a user's insights to a new data or domain. As noted earlier, most systems that are successful at encoding a user's interactions have mechanisms that allow for the reapplication of the interactions within the same system [5, 11, 12, 13]. However, in most analytical environments, analysts often utilize multiple tools simultaneously which renders the use of existing methods inadequate. A more comprehensive and cohesive encoding, recovering, and reusing process is therefore necessary to support the analysts in their natural working environments.

Key Questions to Discuss

Although various user interaction logging technologies exist, we still lack a fundamental understanding of how user interactions can be captured and transformed to insights, and how a visual analytics system can utilize such insights to assist a user in performing future analytical tasks. A number of issues remain open for investigation, and this workshop aims to bring together researchers to examine these issues critically based on their experiences in studying user interactions and provenance capturing. Using the five stages of analytic provenance, these questions can be categorized into:

- Perceive: How is information visually presented to the user that affects the user's reasoning process?
- Capture: What types of user interactions should be captured, and how much semantic information should be included based on a user's task?
- Encode: How should the system store the recorded user interactions? Can the encoded interaction be shared across multiple systems?

- Recover: Based on captured interactions, how can a user's reasoning process be recovered? Can the recovery be done automatically (by a computer)?
- Reuse: How can a visual analytics system apply what it has learned about a user's reasoning process to assist the user in performing future analyses? Can the learned reasoning process be applied to other tasks and other systems?

Expected Participation

We have received a significant amount of interest from diverse groups of researchers in academia, government labs, and industry who have been investigating the relationships between process, interaction, and insight. We therefore expect these participants to bring their expertise in computer graphics, scientific visualization, information visualization, visual analytics, sensemaking, decision making, and HCI to this workshop. With such different backgrounds and interests, we believe that a significant and impactful research agenda can be developed that will be used as a roadmap of future research in the theory and practice of analytic provenance.

References

[1] Thomas, J. and Cook, K. Illuminating the path. National Visualization and Analytics Center, 2004.

[2] Dou, W., Jeong, D.H., Stukes, F., Ribarsky, W., Lipford, H.R., and Chang, R. Recovering reasoning process from user interactions. *IEEE Computer Graphics and Applications* (2009). 29(3): 52–61.

[3] Grabler, F., Agrawala, M., Li, W., Dontcheva, M., Igarashi, T. Generating Photo Manipulation Tutorials by Demonstration. *SIGGRAPH* (2009), 66:1-66:9.

[4] Bavoil, L., Callahan, S., Crossno, P., Freire, J., Scheidegger, C., Silva, C., and Vo, H. Vistrails: enabling interactive multiple-view visualizations. *Proc. IEEE Visualization* (2005), 135-142.

[5] Xiao, L., Gerth, J., and Hanrahan, P. Enhancing visual analysis of network traffic using a knowledge representation. *Proc. IEEE Visual Analytics Science and Technology* (2006), 107–114

[6] Pike, W.A., Stasko, J., Chang, R., and O'Connell, T.A. The science of interaction. *Information Visualization* (2009), 8(1): 263-274

[7] Heer, J., Mackinlay, J., Stolte, C., and Agrawala, M. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics* (2008), 14(6):1189–1196.

[8] Shrinivasan, Y. and van Wijk, J. Supporting the analytical reasoning process in information visualization. *Proc. ACM CHI* (2008), 1237–1246.

[9] Fink, G.A., North, C.L., Endert, A., and Rose, S. Visualizing cyber security: Usable workspaces. *Proc. Visualization for Cyber Security* (2009), 45-56.

[10] Gotz, D. and Zhou, M. Characterizing users' visual analytic activity for insight provenance. *Proc. Visual Analytics Science and Technology* (2008), 123–130.

[11] Jankun-Kelly, T., Ma, K. and Gertz, M. A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics* (2007), 13(2): 357–369.

[12] Garg, S., Nam, J., Ramakrishnan, I., and Mueller, K. Model-driven visual analytics. *Proc. IEEE Visual Analytics Science and Technology* (2008), pp. 19–26.

[13] Kadivar, N., Chen, V., Dunsmuir, D., Lee, E., Qian, C., Dill, J., Shaw, C., Woodbury, R. Capturing and supporting the analysis process. *Proc. IEEE Visual Analytics Science and Technology* (2009), 131-138.