Provenance Data Visualization for Sensemaking

Phong Hai Nguyen

Associate Professor Kai Xu

Professor William Wong

School of Science and Technology
Middlesex University

This dissertation is submitted for the degree of
Doctor of Philosophy

May 2016

Contents

Chapter 1
Introduction

Introduce:

– sensemaking and explain why supporting it is important and challenging

– vis and its capability of supporting sensemaking

1.1 Research Problem and Approach

1.1.1 Problem

Overall, this research aims to *examine how to support users making sense of their problems*.

Introduce analytic provenance and justify why it is a potential approach to support sensemaking.

Summarize the process in which analytic provenance provides support to sensemaking.

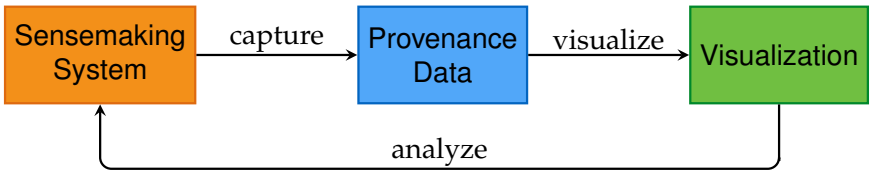


Figure 1.1: A pipeline of supporting sensemaking through analytic provenance. During the process in which a user solves a sensemaking task using a visualization tool, their interactions and discoveries – we refer to both of them as *provenance data* – are captured, visualized and analyzed to provide support back to the sensemaking process.

In this thesis, we focus on the **visualization** part of that process:

How to design interactive visualizations of provenance data to support sensemaking?

1.1.2 Approach

Two types of provenance data:

- user thinking: typically from user notes, high user effort, rich semantics
- user interaction: automatic capture, no user effort, poor semantics

Two types of relationship in the sensemaking process that provenance data may help to reveal:

- temporal: understand the process ('how')
- logical: understand the rationale ('why')

To address the research problem, we take an incremental approach, progressing from the least impact (high user effort to understand the process) towards the most impact (no user effort to understand the rationale).

Two domains to demonstrate:

- intelligence analysis: rigorous sensemaking problems
- everyday sensemaking with web browser: popular, high demands

1.2 Thesis Contributions

Towards the overall goal of supporting users in their sensemaking processes, this thesis contributes

- a compact yet aesthetically pleasing timeline visualization technique that enables users to explore and construct temporal narratives from user annotations.
- a novel timeline visualization technique that enables users to explore more complex temporal narratives by effectively showing both temporal and thematic information.
- a novel application of analytic provenance approach to qualitative data analysis that enables researchers to understand the users' sensemaking processes and a visualization tool to demonstrate its success.
- a light-weight visualization tool (as a browser plugin) system that enables users to see their logical sensemaking processes, to organize the relevant information in such a way that aids their understanding about their problems, and to communicate their findings at different levels of granularity.

1.3 Thesis Outline

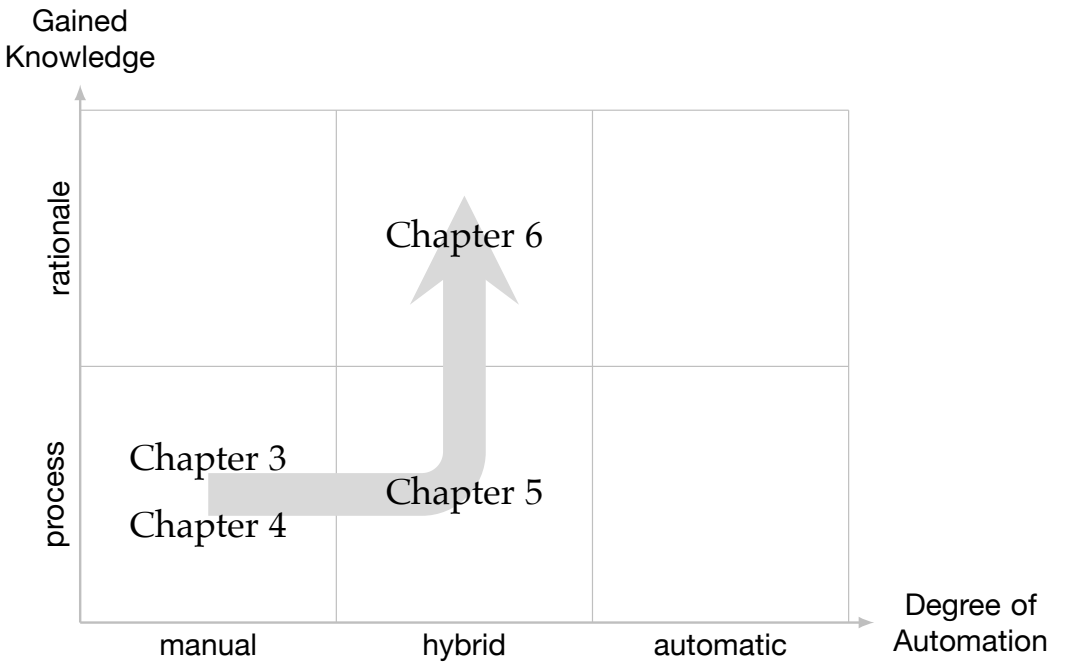


Figure 1.2: Positioning upcoming chapters into the analysis of the degree of automation and the gained knowledge.

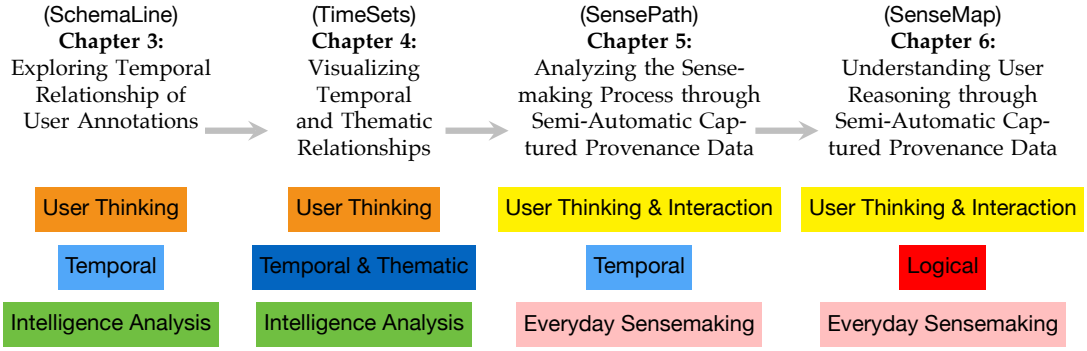


Figure 1.3: Summary of upcoming chapters.

Chapter 2
Literature Review

2.1 Sensemaking

- Overview of sensemaking
    - Pirolli and Card's model
    - Data–frame model

2.2 Visualization and Visual Analytics

- Overview of vis, scivis, infovis, visual analytics
    - Show how they're used to support sensemaking

2.3 Analytic Provenance

- overview of provenance, other types of provenance, then analytic provenance
    - modeling of analytic provenance: 4-level model by Gotz and Zhou and other vis task/action taxonomies
    - pipeline: capture - visualize - reuse and a summary of 'capture' and 'reuse'. the three next sections are details about 'visualize'

2.4 Exploring Temporal Relationship of Provenance Data

Overview of different techniques to visualize temporal data
    Focusing on timeline
    Timeline for sensemaking

2.5 Visualizing Temporal and Thematic Relationship of Provenance Data

Set visualization

- Timelines with set relations

2.6 Understanding User Reasoning through Provenance Data

Recovering user reasoning through interaction logs

- VA systems providing 'reasoning workspace'
- Other non-visualization (data mining) work

Chapter 3

Exploring Temporal Relationship of User Annotations

Timeline visualization is an important tool for sensemaking. It allows analysts to examine information in chronological order and to identify temporal patterns and relationships. However, many existing timeline visualization methods are not designed for the dynamic and iterative nature of the sensemaking process and the various analysis activities it involves. In this chapter, we introduce a novel timeline visualization, SchemaLine, to address these deficiencies.

3.1 Introduction

The importance of timeline visualization in supporting sensemaking.

The current limitation

- no or very simple layout which is cluttered and space-inefficient
- designed for presenting a known story rather than interactively constructing a hidden one

The contribution

- a novel design for an interactive timeline that groups notes into schema determined by the analyst,
- an algorithm to automatically generate a compact and aesthetically pleasing visualization of these schema on the timeline, and
- a set of fluid interactions with the timeline to support the sensemaking activities defined in the Data-Frame model.

3.2 Requirements

[not in the paper yet! List requirements that SchemaLine need to support including the ability to sensemaking activities from Data–frame model]

3.3 Visual Design

Visual representation of individual events and schemas

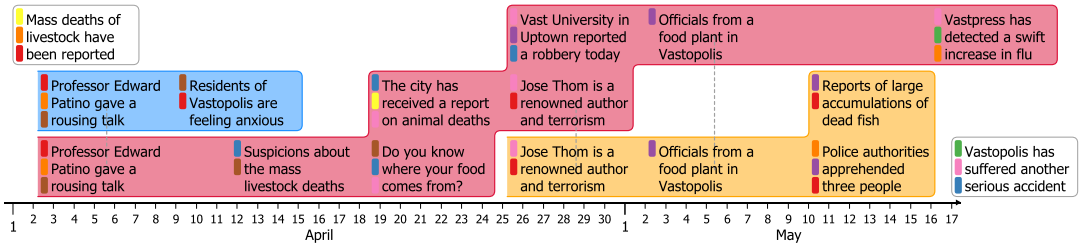


Figure 3.1: SchemaLine: each piece of text is an analyst note, positioned along the time axis at when the event happened. Related notes are linked together to form a "schema" or "frame". There are three frames in this example represented as colored rectilinear paths. Small color-coded rectangles on the left side of notes are "categories".

3.4 Algorithm

Describe the algorithm that produces the compact layout of schemas and events.

3.5 Application

Discuss the integration of SchemaLine into INVISQUE. SchemaLine receives data input as user notes of INVISQUE and the linking between schemas and index-cards.

3.6 Evaluation

A case study with 3 participants (different backgrounds) to use INVISQUE+SchemaLine to solve an intelligence analysis task using VAST Challenge 2011 dataset. Report how they used the tool, how the tool might help them.

3.7 Conclusion

Chapter 4
Visualizing Temporal and Thematic Relationships

The timeline visualization in Chapter 4 cannot show events belonging to multiple sets. In this chapter, we introduce a novel timeline visualization technique, TimeSets, that helps make sense of complex temporal datasets by showing the set relationships among individual events. TimeSets visually groups events that share a topic, such as a place or a person, while preserving their temporal order. It dynamically adjusts the level of detail for each event to suit the amount of information and display estate.

4.1 Introduction

The current limitation

- similarity: use colors or shapes to indicate sets – not powerful
- proximity: not space-efficient
- uniform connectedness: cluttered

The contribution: TimeSets

- clearly shows the events within a set over time and their relationships with other sets;
- dynamically adjusts the level of details of each event to suit the amount of information and display estate;
- uses color gradient backgrounds for events belonging to multiple sets and curved set outlines to emphasize its grouping.

4.2 Design

Visual representation of individual events and sets

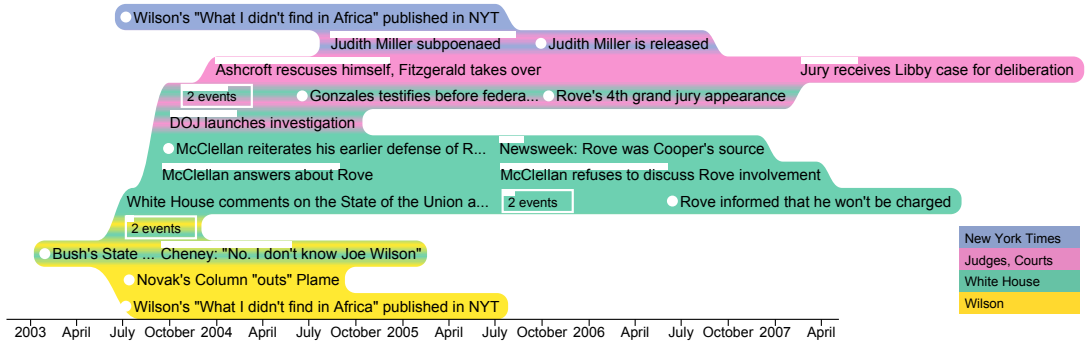


Figure 4.1: TimeSets visualization of the CIA leak case. The timeline contains events that happened from 2002 to 2007, each has a timestamp or an interval, a label, and topics such as “White House”. Events are positioned along the horizontal time axis based on timestamps, and vertically grouped by topics. A time-point event is shown with a white circle to its left, and an interval event with a horizontal bar on top showing its timespan. Each topic has a unique color (see the legend in the bottom right corner), and events shared by two topics have gradient backgrounds, transitioning between the colors of the two topics.

4.3 Layout

Describe the layout algorithm.

4.4 Evaluation

A controlled experiment with 30 participants to compare TimeSets and the state-of-the-art set visualization technique – KelpFusion.

4.5 Case Study 1: Publication Data

Shows an application of TimeSets to publication data of 200 articles and 8 sets.

4.6 Case Study 2: VAST Challenge 2014

Shows an application of TimeSets to tweets data.

Also use TimeSets to show user findings [not done yet! currently, findings are shown as a node-link diagram.]

4.7 Conclusion

Chapter 5

Analyzing the Sensemaking Process through Semi-Automatic Captured Provenance Data

In this chapter, we consider a specific sensemaking task: *a qualitative study to understand user's sensemaking process*. This is an essential task because understanding the sensemaking process allows building effective visual analytics tools to make sense of large and complex datasets. Currently, it is often a manual and time-consuming undertaking to comprehend this: researchers collect observation data, transcribe screen capture videos and think-aloud recordings, identify recurring patterns, and eventually abstract the sensemaking process into a general model. We propose a general approach to facilitate such a qualitative analysis process, and introduce a prototype, *SensePath*, to demonstrate the application of this approach with a focus on *browser-based online sensemaking*.

5.1 Introduction

Current limitation: transcription and coding are very time-consuming.

Contribution

1. A general approach combining the strength of analytic provenance and visual analytics to understand user's sensemaking process. This approach can be potentially applied to other qualitative research in HCI beyond sensemaking.
2. A qualitative study and a participatory design session to understand characteristics of qualitative research on sensemaking.
3. A visual analytics tool SensePath to demonstrate the general approach. It supports the transcription and coding of the observation data of online sensemaking tasks.

4. A qualitative user evaluation that demonstrated the effectiveness of the general approach and the tool SensePath.

5.2 Design Research and Requirements

Discuss how we approach the problem and elicit requirements to support.

5.3 Interface Design

Describe the tool interface

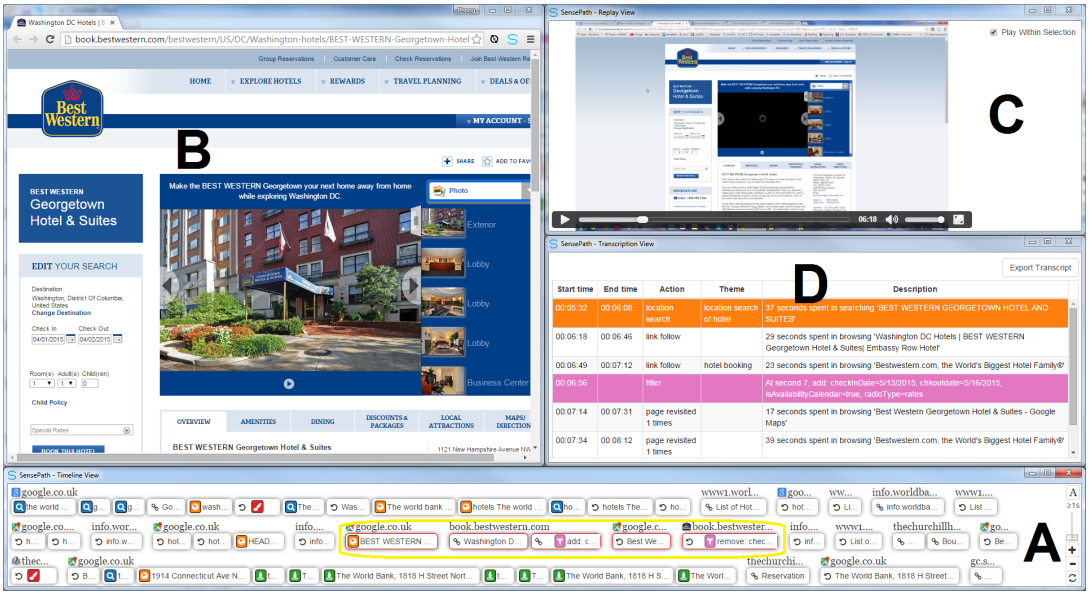


Figure 5.1: Four linked views of SensePath. **A:** The *timeline* view shows all captured sensemaking *actions* in temporal order. **B:** The *browser* view displays the web page where an action was performed. **C:** The *replay* view shows the screen capture video and can automatically jump to the starting time of an action when it is selected in another view. **D:** The *transcription* view displays detailed information of selected actions (the highlighted ones in the timeline view).

5.4 Evaluation

We conducted two user-centered evaluations: the first one is to understand how SensePath is used by an experienced qualitative researcher and to identify opportunities for improvement, and the second one is to discover whether SensePath has any advantages compared to a traditional method.

5.5 Conclusion

Chapter 6

Understanding User Reasoning through Semi-Automatic Captured Provenance Data

Very often, users get lost when solving a complicated task using a big dataset over a long period of exploration and analysis. They may forget what they have done, are not aware of where they are in the context of the overall task, and do not know where to continue. In this chapter, we introduce a tool, *SenseMap*, to address these issues in the context of *browser-based online sensemaking*.

6.1 Introduction

Limitation: Existing approach – graphical browser history – only provides a static overview of the browsing process

Contribution:

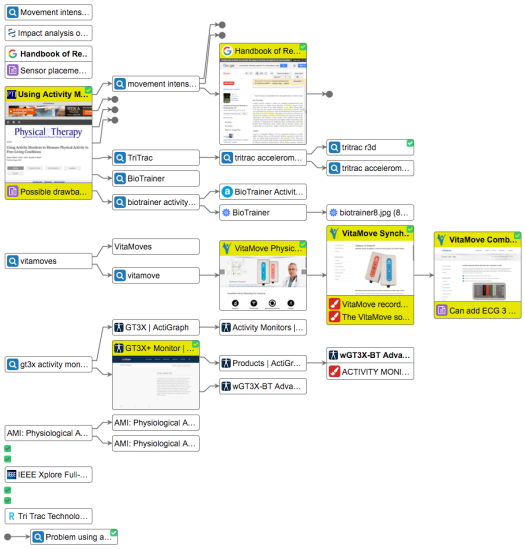
1. A user study exploring how users search, manage and synthesize online information for their daily work activities; and a series of workshops followed up to generate design questions and formulate solutions.
2. A visual analytics tool SenseMap supporting browser-based online sensemaking addressing all the elicited requirements.
3. A user evaluation exploring how SenseMap is used in a naturalistic work setting and a process model derived from the data analysis.

6.2 Design Research and Requirements

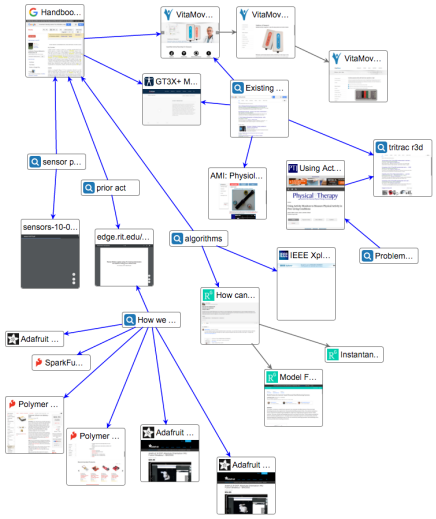
Describe the study to elicit requirements

6.3 Interface Design

Describe the tool addressing all the requirements



(a) History Map: visualizes captured user sensemaking actions to provide the overview of the sensemaking process.



(b) Knowledge Map: curates and makes sense of the most relevant information to the task.

Figure 6.1: SenseMap interface.

6.4 Evaluation

We conducted a user-centered evaluation of SenseMap in order to:

- Evaluate its effectiveness in providing the desired support for searching, managing and synthesizing information through our collect, curate, communicate process model;
- Identify significant features relating to user behaviors, interactions and outcomes in the use of this new tool; and
- Describe any processes that these behaviors, interactions and outcomes might infer.

6.5 Conclusion

References