1 Literature Review

1.1 Sensemaking

Overview of sensemaking

1.1.1 Pirolli and Card's Model

Pirolli and Card [25] describe sensemaking as a process consisting of two major loops: the foraging loop and the sensemaking loop (Figure 1.1).

1.1.2 Data–Frame Model

Klein et al. [16] defines sensemaking as a deliberate effort to understand an event, starting when a person realizes the gap of their current understanding of that event. These sensemaking activities are summarized in Figure 1.2.

1.2 Visualization and Visual Analytics

- defines visualization [22]
- how vis helps support analysis [7]
- vis reference model [7]
- difference between scivis and infovis
- define visual analytics [15, 35]
- visual analytics process model (Figure 1.3)

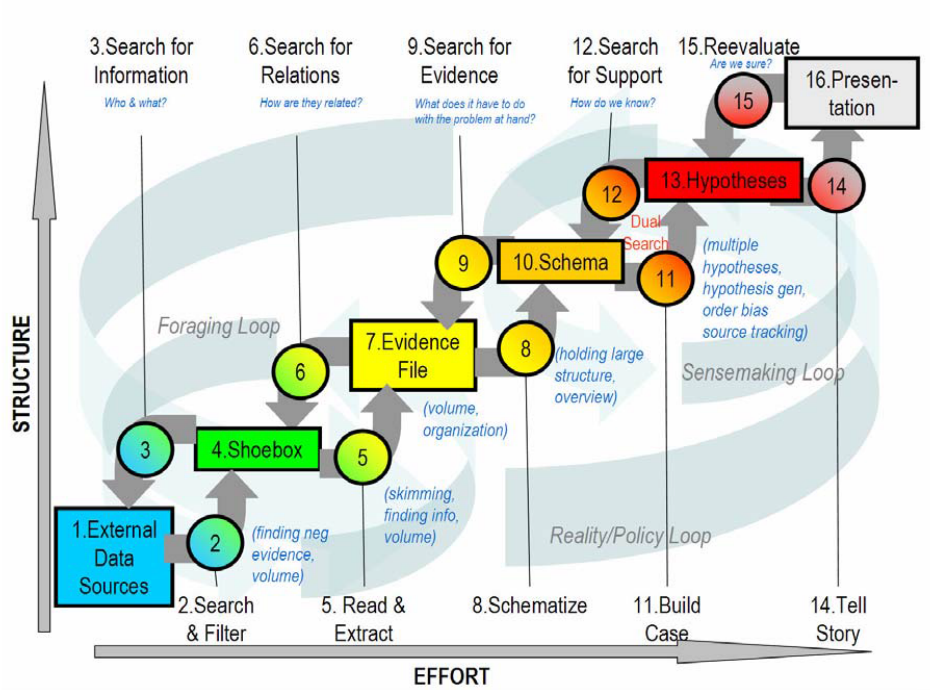


Figure 1.1: A sensemaking model proposed by Pirolli and Card. *Source: [25]*.

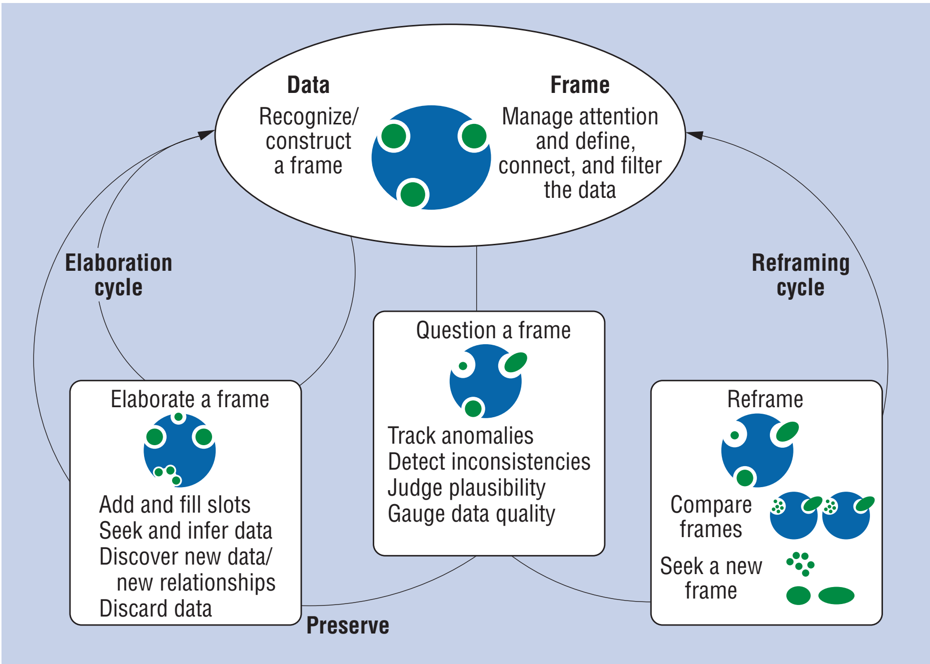


Figure 1.2: The data–frame model. *Source: [16]*.

some components of visual analytics (from William's slides) (they're similar to the new definition of visual analytics in Mastering the Information Age "Visual

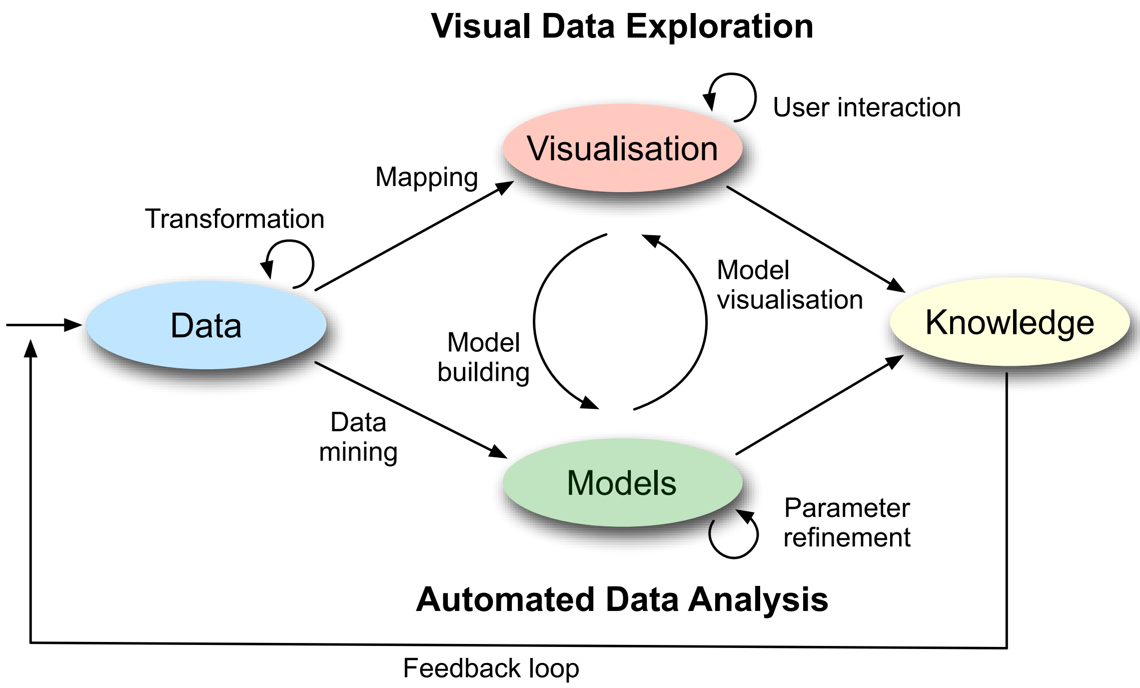


Figure 1.3: The data–frame model. *Source: [16]*.

analytics combines automated analysis with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets.")

- data analysis/data mining
- interactive dynamics (Heer and Shneiderman)
- assembly of evidence, argumentation (toulmin, wigmore)

1.3 Information Design Principles

- encoding channel ranking (chap 5, 6 Munzner book)
- Focus+Context
- Gestalt Principles
- Principle of Visual Affordance
- Tufte: ink-ratio
- Ben Shneiderman's influential mantra of Overview First, Zoom and Filter, Details on Demand

Several interface design guidelines: Schneiderman (eight golden rules), Norman (design principles) and Nielsen's (ten usability heuristics)

1.4 Analytic Provenance

Overview of provenance – provenance in ordinary context, give examples (painting)

Some types of provenance: database, workflow, semantic web (see the previous review of analytic provenance)

Then analytic provenance: definition from [23], our agenda [42]

Modeling of analytic provenance: 4-level model by Gotz and Zhou and other vis task/action taxonomies such as classic Shneiderman's task by data type [29], Amar's low-level [1], Amar's high-level [2], more recent Bhemer's typology what-why-how [5].

Pipeline: capture - visualize - use and a summary of 'capture' and 'use' (except for directly supporting sensemaking)

1.4.1 Capture

combine previous reviews from the draft survey of analytic provenance, research agenda, and capture in the context of online sensenmaking the last paper

1.4.2 Use

work from the draft survey of analytic provenance: reusing the performed analysis (workflow), making tutorial, presentation

1.5 Visualizing Temporal Relationship of Sensemaking

Visualization of Temporal Events

- sequential timeline as comic strips
  - overview: [19]
  - examples: [10], DIVA [36]
- continuous timeline
  - one of the earliest classic examples: Chart of Biography [28], Gantt chart [8]

  - (events as glyphs) early visualization: LifeLines [26, 27] and its extensions LifeLines2 [37, 38], Similan [40], LifeFlow [39], for interval events: EventFlow [20] [describe what's new in them]
  - (events as user annotations) POLESTAR [24], HARVEST [11], Jigsaw [9, 33], and nSpace2 Sandbox [31, 41] – how do they support sensemaking?
  - some online services

- Storyline: show temporal dynamics of social interactions [18, 21, 34]
- ThemeRiver: show thematic changes in large document collections – aggregate a number of events into intervals, kind of smooth histogram [6, 13]

Viusialization of Event Relationships

- drawing edges between related events: tmViewer [17]
- mutliple facets, views with different scales: Continuum [3]
- time-dependent and time-independent (how?): SemaTime [32]

1.6 Visualizing User Reasoning in Sensemaking

Visualization of Non-Linear Sensemaking Process Visualize the branching history

- typically, use tree; examples: [14], VisTrails: [4], or general graph if 'undo' action is explicited drawn
- overview of tree/graph visualizations?
- 'behavior graph' in Tableau
- encode time into graph: color coded [4], edge length [30]

Visualization of Analytical Reasoning

- typically, use graph; examples: [30], Scalable Reasoning System, sandbox
- more formal visual arguments? that support toulmin, wigmore

I think set visualizations will be described in the related work of TimeSets and qualtitative research in SensePath

2 Exploring Temporal Relationship through User Annotations

As the first step toward supporting sensemaking using analytic provenance, this chapter examines how to enable users to explore the temporal relationship hidden in the sensemaking task through the annotations they made. *Timeline* is a simple yet powerful technique to visualize time-oriented data, enabling users to examine information chronologically and 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 issues.

2.1 Introduction

Explain the importance of timeline visualization in supporting sensemaking.

Limitation of existing timeline visualization techniques:

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

SchemaLine contributes

- A novel design for an interactive timeline that groups notes into schemas determined by the user

- An algorithm to automatically generate a compact and aesthetically pleasing visualization of these schema on the timeline
- A set of fluid interactions with the timeline to support the sensemaking activities defined in the Data-Frame model

2.2 Design Requirements

Describe how the following requirements are elicited. Define 'event' and 'schema'.

1. **Automatic layout**. Provide an automatic layout to leverage users from manually arrange the events.
2. **Natural flow**. Easy to follow events within the same schema chronologically.
3. **Data–frame activities**. Enable users to perform sensemaking activities described in the Data–frame model through fluid interactions.

2.3 Visual Design

2.3.1 Event

Figure 2.1: An event with three categories. The event is represented as a rounded rectangle with summarized content inside. On the left side of the text, small color-coded rectangles indicate its categories. The event is positioned along the time axis at which it happens – 8th of May.

2.3.2 Schema

Putting it all together, Figure 2.3 shows an example of a complete SchemaLine. The algorithm to produce this visualization is described in Section 2.4.

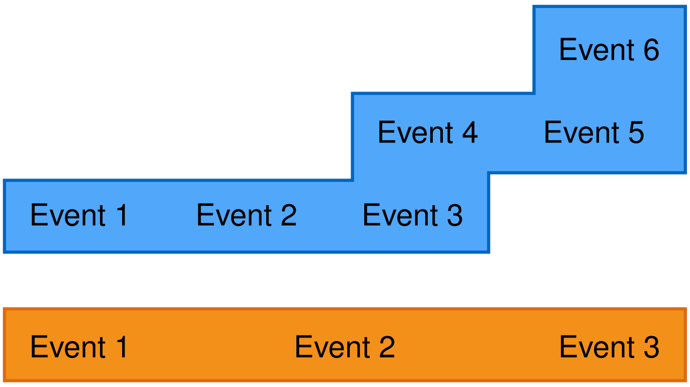


Figure 2.2: A schema as a colored stripe covering its events. Bottom: a simple rectangle connecting events that can display in the same row. Top: a rectilinear path connecting events that need to locate in different rows. Colored background distinguishes different schemas.

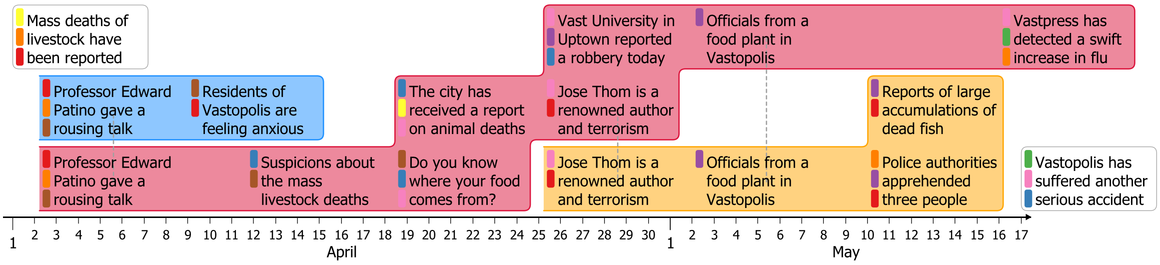


Figure 2.3: SchemaLine visualization of user annotations. Related notes are connected to form schemas.

2.3.3 Interaction

Describe how interaction in SchemaLine supports different sensemaking activities in the Data–frame model (Requirement 3).

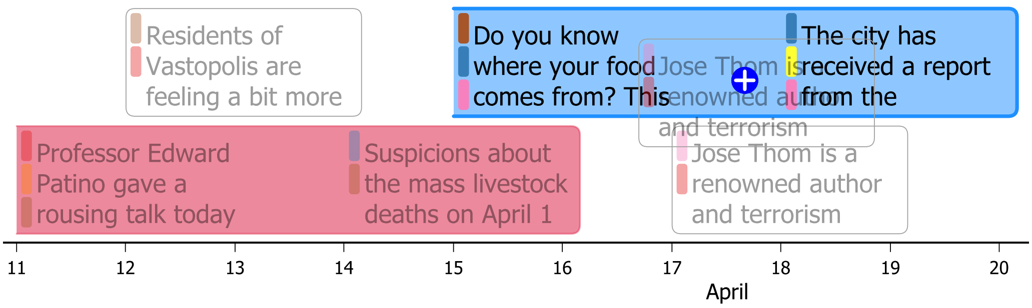


Figure 2.4: Elaborating a frame. Dragging and dropping an event onto the blue stripe to add it to the frame.

2.4 Algorithm

The algorithm consists of two parts. First, the layout of schemas is generated; and then their outlines are computed based on the layout information.

2.4.1 SchemaLine Layout

This algorithm is illustrated in Figure 2.5.

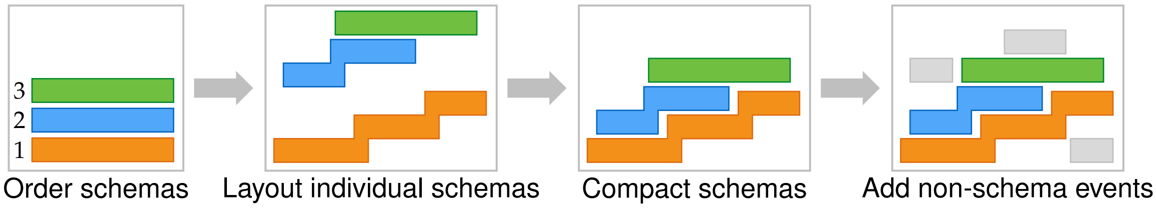


Figure 2.5: SchemaLine layout algorithm. First, the order of all schemas is computed. Second, the layout of each schema is generated independently. Third, schemas are compacted following the computed order. Finally, events that do not belong to any schemas are added.

Figure 2.6 shows a step-by-step illustration of the algorithm in the second step for a timeline with four events.

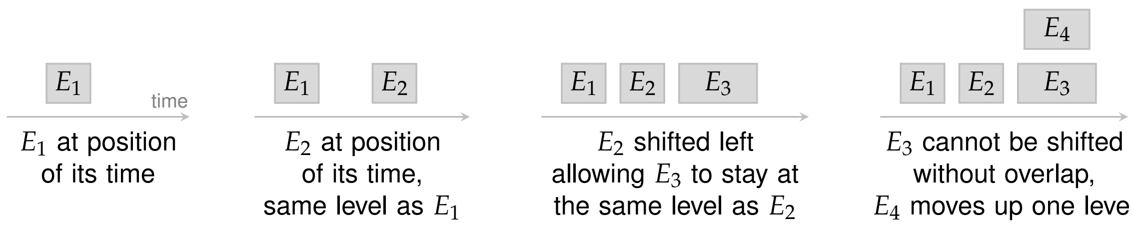
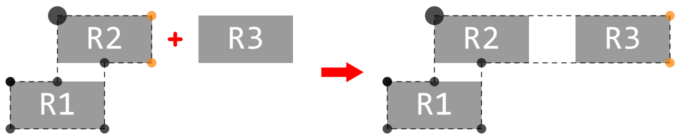


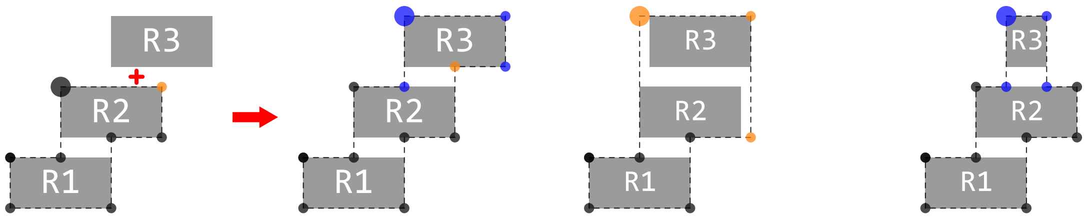
Figure 2.6: Schema layout algorithm. Four events E_1, E_2, E_3, E_4 will be added to their schema chronologically. E_1 is positioned at its accurate event time. E_2 can locate in the same level as E_1 because it does not intersect with E_1. E_3 intersects with E_2 but the intersection width is small enough for E_2 to shift to the left without intersecting with E_1, enabling E_3 locate in the same level as well. However, E_4 needs to move up one level because it is impossible to avoid overlapping while shifting previous events.

2.4.2 Schema Outline

In this section, we describe an algorithm to produce a polygonal outline covering all the event rectangles of a schema.



(a) R3 is on the right side of the path.



(b) R3 is on top of the path.

(c) Left of R3 is a bit greater than left of R2.

(d) Right of R3 is shorter than right of R2.

Figure 2.7: Four cases when a new rectangle R3 is appended to the polygonal path. A big circle indicates the pivot vertex of the path (top-left corner of the last rectangle). Orange circles indicate updated vertices, and blue circles indicate newly added vertices of the polygonal path.

2.5 Application

We integrate SchemaLine into INVISQUE – INteractive VIsual Search and QUery Environment – designed for interactive exploration of text documents.

2.6 Evaluation

- Goal: explore how SchemaLine is used in a sensemaking task
- Method: qualitative study
- Task: Mini Challenge 3 of the VAST Challenge 2011, which requires participants to identify any potential criminal activities from a given dataset
- Participants: 3 with different backgrounds
- Procedure: 10-minute training → one-hour main task → semi-structured interview
- Report: describe three case studies separately, followed by a general discussion of all three; focusing on how their sensemaking activities map to activities in Data–frame model and performed by interaction with SchemaLine

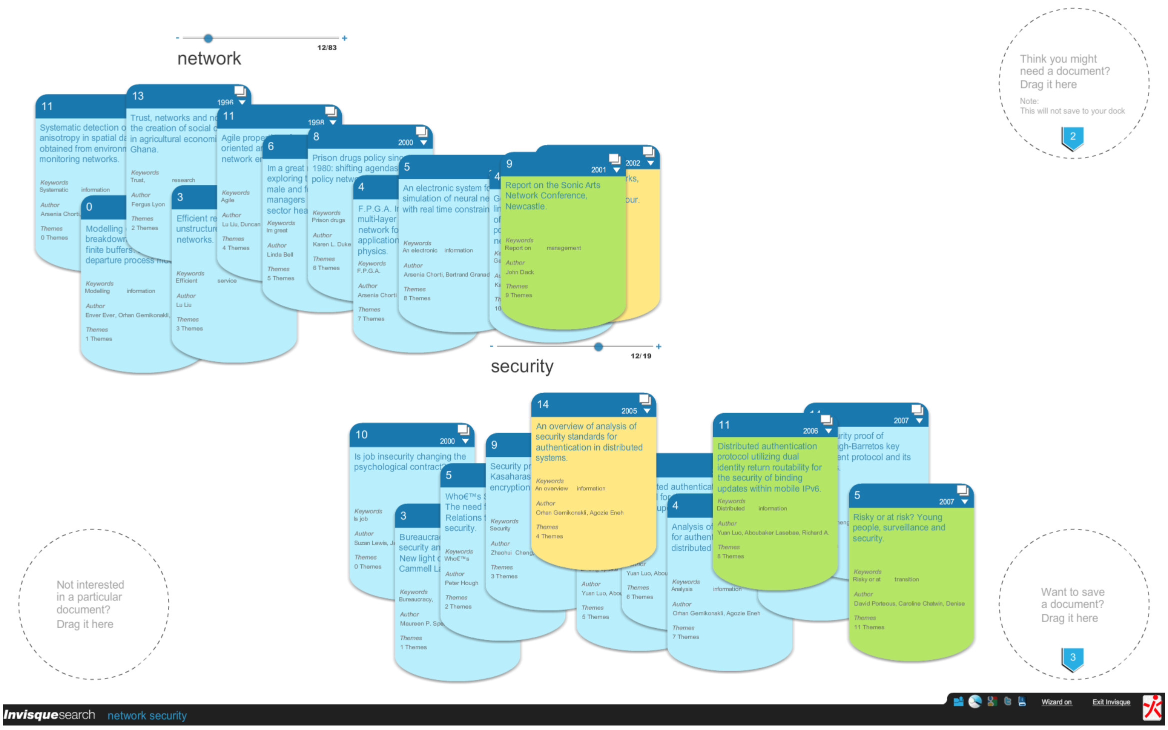


Figure 2.8: INVISQUE interface. It shows two sets of search results for "network" and "security" from a publication dataset. Each article is displayed as an index-card with citation count at the top left corner, publication year at the top right corner; and title, keywords and authors at the center.

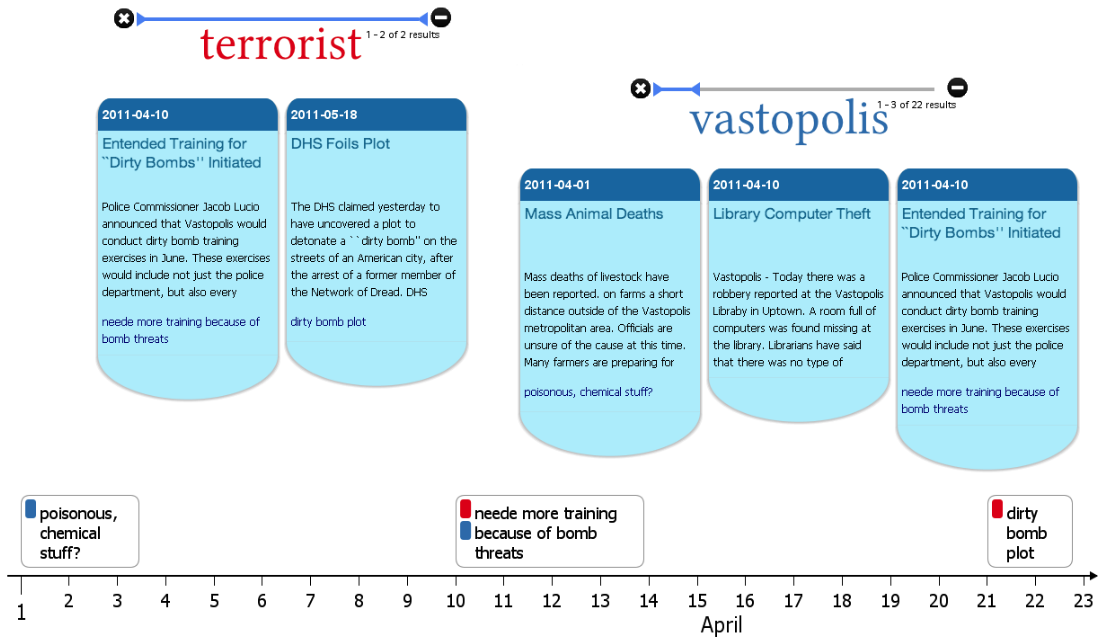


Figure 2.9: INVISQUE with SchemaLine at the bottom. The timeline consists of three user annotations with the color coded categories indicating keywords that were searched for.

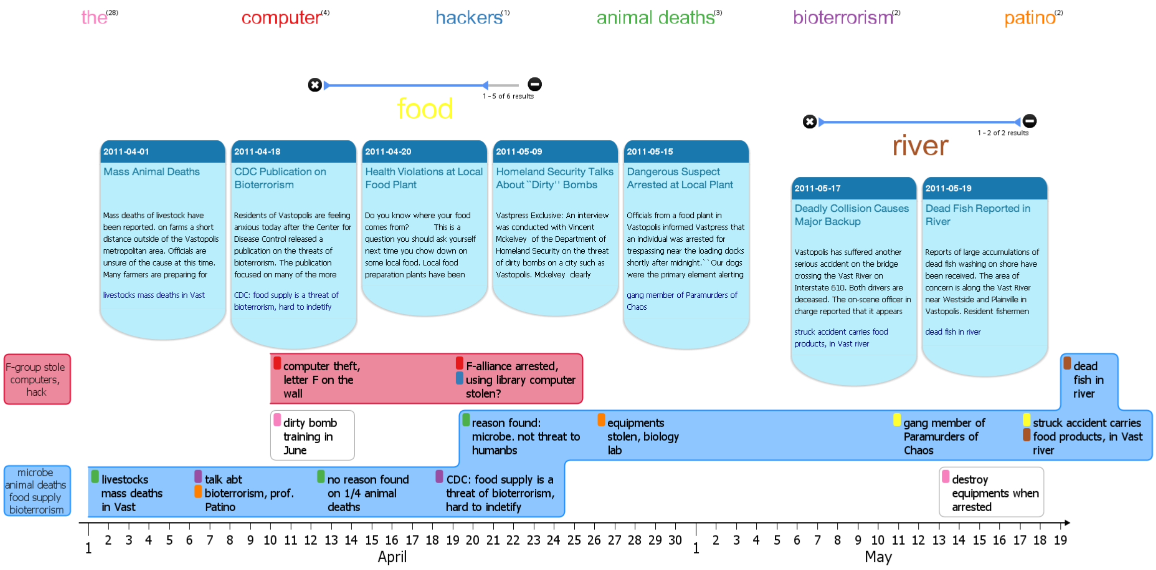


Figure 2.10: Final screen of participant **P2**. Top: a trail of his keyword searches, collapsed after being read. Middle: search results in index-card metaphor. Bottom: two schemas containing notes as supporting evidence of criminal activities he found.

3 Exploring Complex Temporal Relationship

The SchemaLine technique described in Chapter 3 cannot show events belonging to multiple sets limiting users from exploring complex temporal relationship. In this chapter, we introduce a novel technique – TimeSets – to address that issue.

3.1 Introduction

Limitation of existing timeline visualization techniques when showing set relations:

- 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.

3.2 Visual Design

3.2.1 Event

Figure 3.1 shows examples these different visual representation of events.

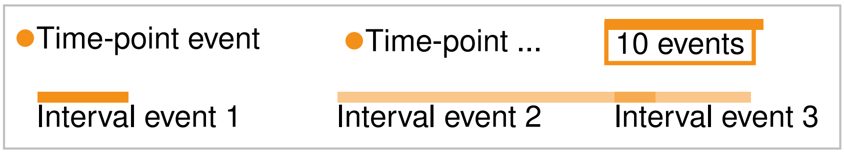


Figure 3.1: Visual representations of events. Top row, left to right: a complete time-point event, a trimmed time-point event, and a aggregate of 10 events. Bottom row, left to right: an interval event, and two overlapping interval events.

3.2.2 Set

3.2.2.1 Design Overview

We use two most effective Gestalt's principles of grouping in our design: *proximity* and *uniform connectedness*. Events belonging to the same set are located close together, and the background of an entire set is colored to make its events visually connected.

Spatial grouping is achieved through vertical positioning because the horizontal position of each event is already fixed by its temporal information. Figure 3.2 shows an example of a layering for three sets.

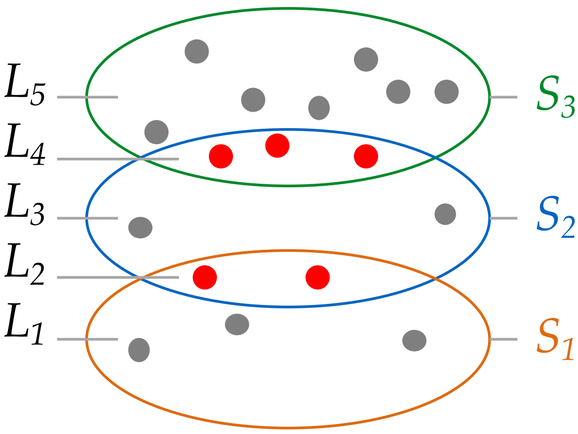
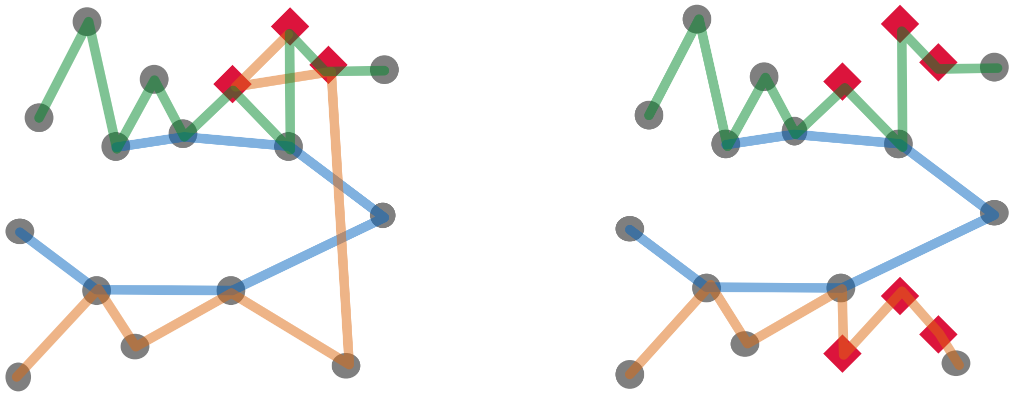


Figure 3.2: Layering for three sets S_1, S_2, and S_3. L_2 includes events shared by S_1 and S_2, and L_4 includes events shared by S_2 and S_3. Shared events are shown in red. S_2 consists of events in three layers L_2, L_3, and L_4.

Discuss design choices for visualizing shared events between two non-neighboring sets (Figure 3.3)



(a) Shared events are only located in the green set, and connected to the orange set.

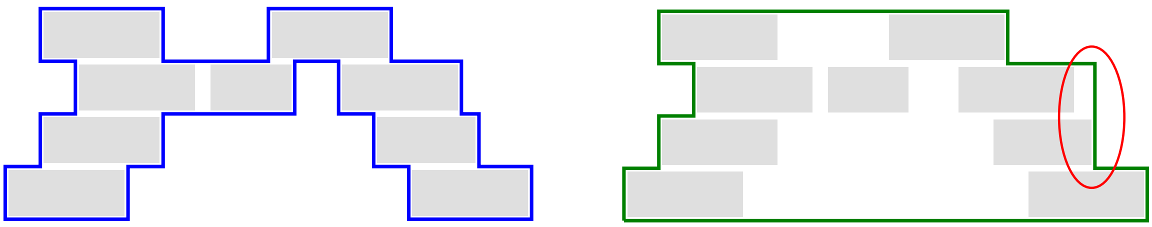
(b) Shared events are duplicated in both green and orange sets.

Figure 3.3: Visualization of shared events (red squares) between two non-neighboring sets.

In subsequent sections, we discuss the detail of the set visualization algorithm, which consists of two main steps: generating set shapes, and then coloring them.

3.2.2.2 Shape Generation

This algorithm takes as input a list of bounding-boxes of the set's events, and generates a closed-curve containing all these boxes.



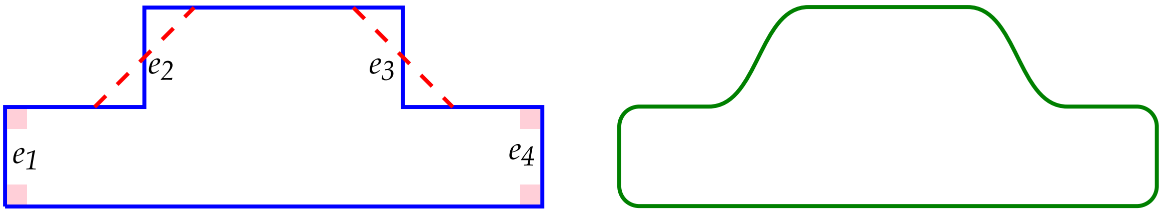
(a) The original rectilinear shape generated by a scan-line algorithm.

(b) The simplified shape by flattening and removing jags (red eclipse).

Figure 3.4: Rectilinear shape generation.

3.2.2.3 Set Coloring

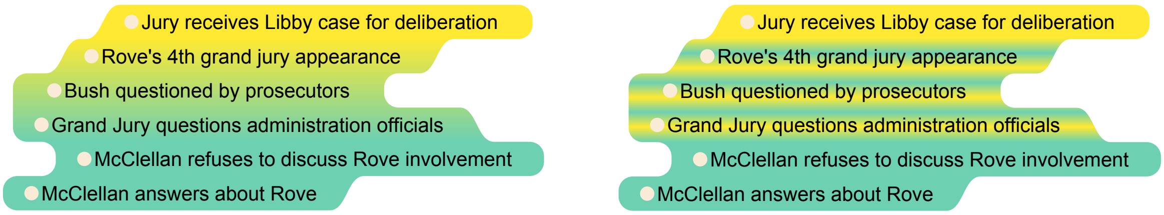
Each set is filled with a color selected from Qualitative Set 2 of ColorBrewer [12] to make them easily distinguishable. And the background of the intersection between two sets is shown as the color transitioned between the two set colors (Figure 3.6).



(a) Vertical segments e_2 and e_3 are converted to diagonal ones (dashed lines).

(b) Squared corners are replaced by quadrant arcs. e_2 and e_3 are further smoothened by Bézier curves.

Figure 3.5: Shape smoothening by reducing the degree of line bends.



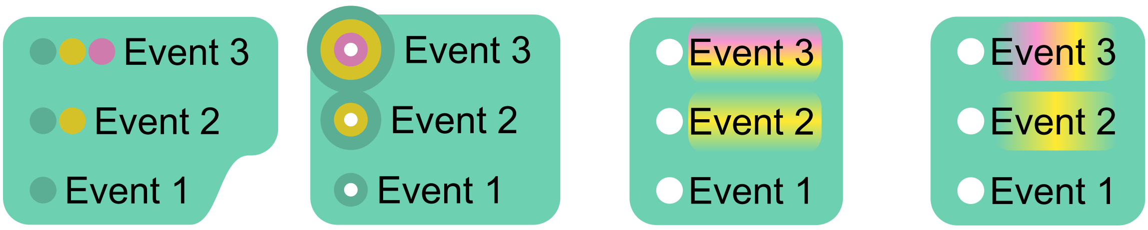
(a) Intersection shown as a single color gradient.

(b) Intersection shown as multiple color gradients.

Figure 3.6: Color gradient technique to encode set intersection. The gradient area shows three shared events between the yellow and green sets.

3.2.2.4 Multiple-set Events

Four design options are proposed to visualize the full set memberships of events.



(a) Each circle represents a set.

(b) Each ring represents a set.

(c) Vertical gradient: each color represents a set.

(d) Horizontal gradient: each color represents a set.

Figure 3.7: Visual representation of multiple-set events. Event 1 is single-set (green). Event 2 is double-set (green, yellow). Event 3 is triple-set (green, yellow, pink).

Putting it all together, Figure 2.3 shows an example of a complete SchemaLine. The algorithm to produce this visualization is described in Section 2.4.

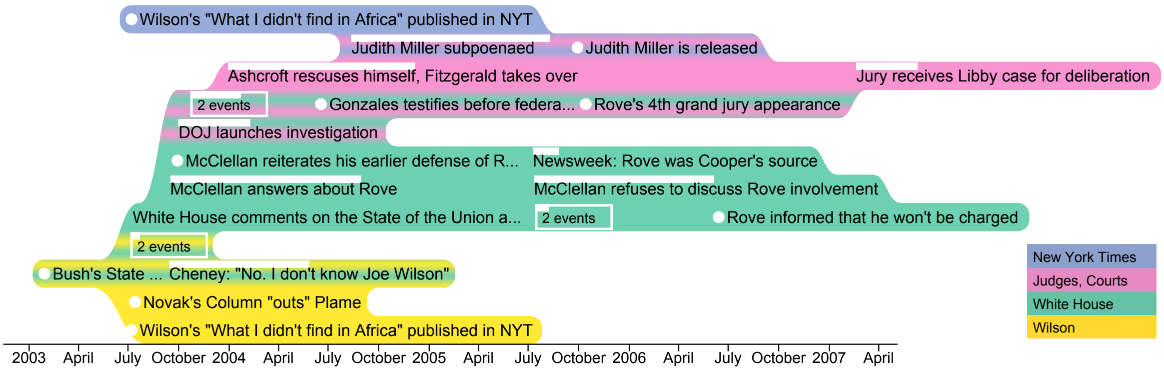


Figure 3.8: TimeSets visualization of the CIA leak case. The timeline contains events that happened from 2002 to 2007, each includes a timestamp, a label, and topics such as “White House”. Events are positioned along the horizontal time axis based on their temporal values, and vertically grouped by colored topics (see the color legend in the bottom right corner).

3.2.3 Interaction

Describe interaction including zoom, pan, mouse hover, filter.

3.3 Layout

3.3.1 Sets Ordering

The same as the schema ordering algorithm in SchemaLine

3.3.2 Layer Layout

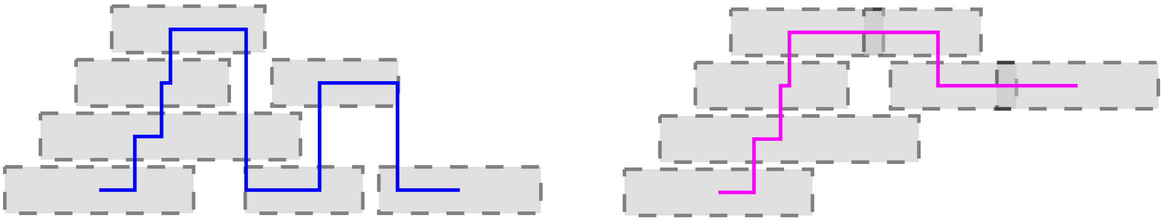
This algorithm positions all the events within a layer. We propose two algorithms: one is optimized to display as much information as possible (*completeness*), and the other one is optimized for the convenience to follow events (*traceability*).

3.3.3 Layers Compacting

Compact layers together leaving no gap rows among layers.

3.3.4 Layers Balancing

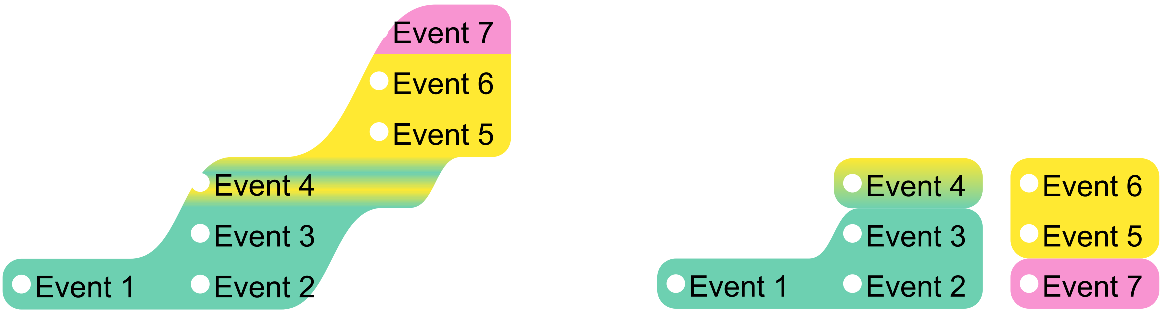
This last step ensures that all layers have similar levels of detail; i.e., avoiding layers with many complete events and other layers with many aggregated events.



(a) Completeness algorithm: $\theta = 1$, $\gamma = 5/3$.

(b) Traceability algorithm ($t_{min} = 0.5$, $r_{max} = 1$): $\theta = 6/7$, $\gamma = 2/3$.

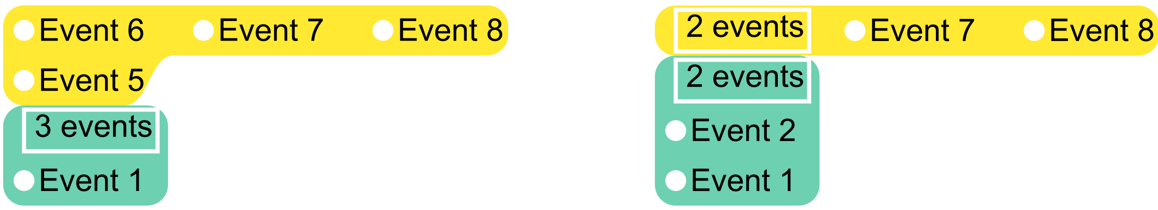
Figure 3.9: Layer layout algorithms. Each rectangle represents an event. The line connecting centers of rectangles illustrates the traceability.



(a) Before: consumed height = 6 rows.

(b) After: consumed height = 3 rows.

Figure 3.10: Layers compacting.



(a) Before: $\theta_{green} = 0.25$, $\theta_{yellow} = 1$.

(b) After: $\theta_{green} = \theta_{yellow} = 0.5$.

Figure 3.11: Layers balancing.

3.4 Evaluation

- Goal: compare task performance between TimeSets and a relevant state-of-the-art visualization technique (KelpFusion)
- Method: lab controlled experiment, within-subject
- Task: 5 different tasks related to temporal and set relations using 4 generated datasets
- Participants: 30, each answers 2 x 5 x 4 = 40 questions

- Procedure: introduction → practice → main task → subjective questionnaire
- Report: RM-ANOVA significant tests to analyze performance data and Fisher's exact tests to analyze user ratings

3.5 Case Study 1: Publication Data

Shows an application of TimeSets to publication data of 200 articles and 8 sets. Discuss interesting findings with TimeSets.

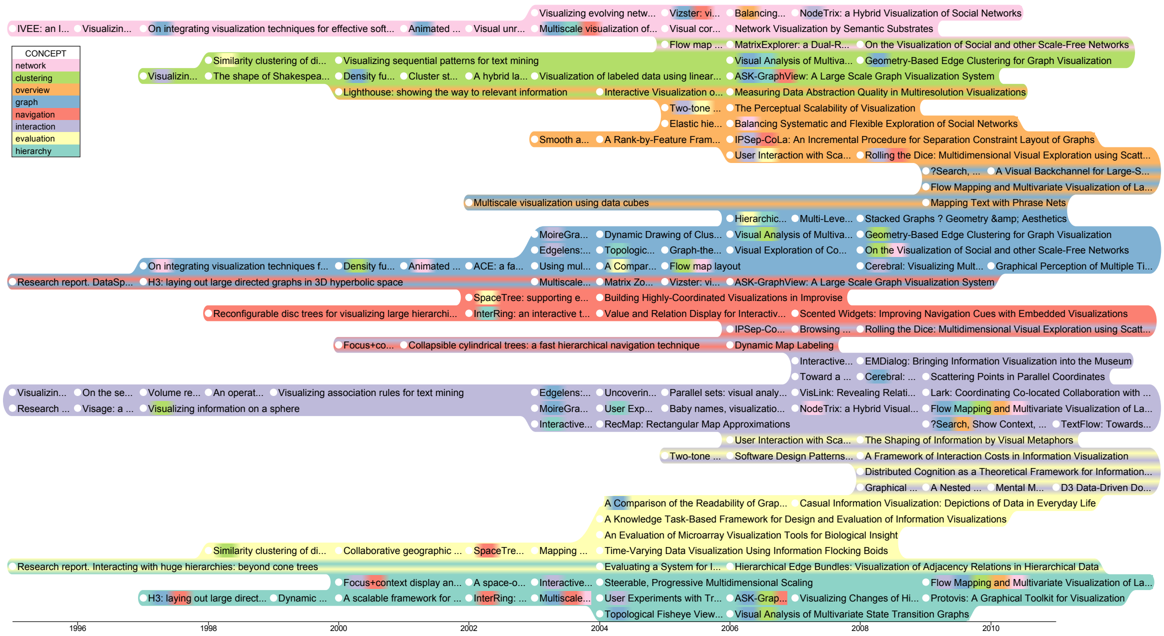


Figure 3.12: TimeSets visualization of publication data. It shows 200 articles with the most citations in the IEEE InfoVis conference from 1995 to 2013. These articles are categorized based on their concepts (see the legend in the top left hand corner).

3.6 Case Study 2: VAST Challenge 2014

Shows an application of TimeSets to intelligence analysis. TimeSets is used to show both tweets ([this is the color blending version of TimeSets, not color gradient – should we describe the color blending one?], Figure 3.13) and findings ([missing figure – currently, findings are shown as a node-link diagram])


sofitees: @carlyscoffee best coffee in Abila
CentralBulletin: expansion of evacuation perimeter in Abila fire #Abila #CentralBulletin
NewsOnlineToday: Trapped resident rescued in Abila fire. #abila #newsOnline
AbilaFireDept: Abila Fire Department trucks 12 & 30 have arrived?at the Dancing Dolphin fire to relieve trucks 41 and 57 #AFD?
CentralBulletin: building fire reported in east Abila police and fire dispatched #Abila #CentralBulletin
Viktor-E: Viktor-E performing at Abila coffee houses every Friday this month: Viktor-E.kronos/schedule
AbilaFireDept: Abila Fire Department has dispatched two fire trucks to the Dancing?Dolphin apartments at?Achilleos & Madeg #AFD?
CentralBulletin: Dancing Dolphin apartment complex identified as loca...
InternationalNews: Neighborhood evacuations in Dancing Dolphin Fire #Abila #intNews
KronosStar: Abila Fire Department has dispatched 2 units to a possible fire ...
NewsOnlineToday: Additional residents being evacuated in Dancing Dolphin fire in Abila, #newsOnline
NewsOnlineToday: Nearby residents being evacuated in Dancing Dolphin fire in Abila. #NewsOnline
AbilaFireDept: Abila Fire Department fire truck #23 has arrived at Dancing Dolphin apartment #AFD
HomelandIlluminations: Residents trickling out of Dancing Dolphin. #HI
AbilaFireDept: Abila Fire Department on the scene of Dancing Dolphin fire at Achilleos & Madeg assessing the extent of the fire #AFD
HomelandIlluminations: Police arrived at Dancing Dolphin fire. #HI
HomelandIlluminations: Smoke from first two floors of Dancing Dolphin apartments. Fire officials on site. #HI
InternationalNews: Continuing fire at the Dancing Dolphin apartment #Abila #IntNews
AbilaPost: The location of the fire has been identified as the Dancing Dolphin Apartment complex #AbilaPost
InternationalNews: Fire at the Dancing Dolphin apartment complex. #Abila #IntNews
KronosStar: Abila Police Department has sent two units to the fire at the Dancing Dolphin apartments. #KronosStar
AbilaPoliceDepartment: Abila Police Department urges any witnesses of the hit and run incident on Schaber between Eira & Ede
AbilaPoliceDepartment: Abila Police Department issues description of hit and run vehicle: a black van with partial license plate L82
AbilaPoliceDepartment: Abila Police Department releases a partial license plate number for hit and run vehicle L829 #APD
AbilaPoliceDepartment: Abila Police Department reports a bicyclist has been struck on the 500 block of Schaber between Eira and Edes
AbilaPoliceDepartment: Abila Police Department reports a car hit on Souliou street #APD
06:45
07 PM
07:15
07:30
07:45
08 PM
08:15
60
40
20
0
05 PM
05:30
06 PM
06:30
07 PM
07:30
08 PM
08:30
09 PM
09:30
positive
negative
neutral
RT
search
PERSON
Hide
Sylvia Marek (0/43)
Lucio Jakab (0/34)
Dr. Newman (0/33)
Sylvia (0/13)
Dr Audrey McConnell Newman (0/12)
Bert Hanns (0/9)
ORGANIZATION
Hide
Abila Police Department (6/19)
Abila Fire Department (5/25)
POK (0/73)
FACILITY
Hide
Dancing Dolphin (14/38)
Dancing Dolphin Apartment (1/9)
Achilleos & Madeg (1/3)
CITY
Hide
Abila (12/56)
Kronos (1/10)
Souliou (1/1)

Figure 3.13

4 Analyzing the Sensemaking Process through Semi-Automatic Captured Provenance Data

Chapter 2 and Chapter 3 enable users to explore complex temporal relationship of sensemaking through interactive timeline visualizations of their annotations. These annotations may contain high-level thinking of the users, facilitating them to understand and exploit effectively. However, making detailed and frequent notes require a huge amount of time and effort from the users. Also, it may distract them from their sensemaking tasks. In this chapter, we capture user sensemaking actions automatically and enable users to analyze both manual annotations and automatic actions through interactive visualizations.

4.1 Introduction

We consider a specific sensemaking task: *a qualitative study to understand user's sensemaking process*. 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.

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.

4.2 Approach and Requirements

4.2.1 Approach

We conducted two sets of observations to explore the characteristics of qualitative analysis of sensemaking activities.

4.2.2 Requirements

1. **Thematic analysis support**
2. ***Transcription* and *Coding* efficiency**
3. **Existing workflow integration**
4. **Non-intrusiveness**
5. **Scalability**
6. **Lightweight**

4.3 Interface Design

4.3.1 Approach

The desgn process started with a close examination of steps we want to support, identify what needs to be done to meet the requirements.

1. **Seeing the actions before and after the current one**.

2. **Seeing what a participant was looking at.**
3. **Understanding what a participant was thinking.**

4.3.2 Overview

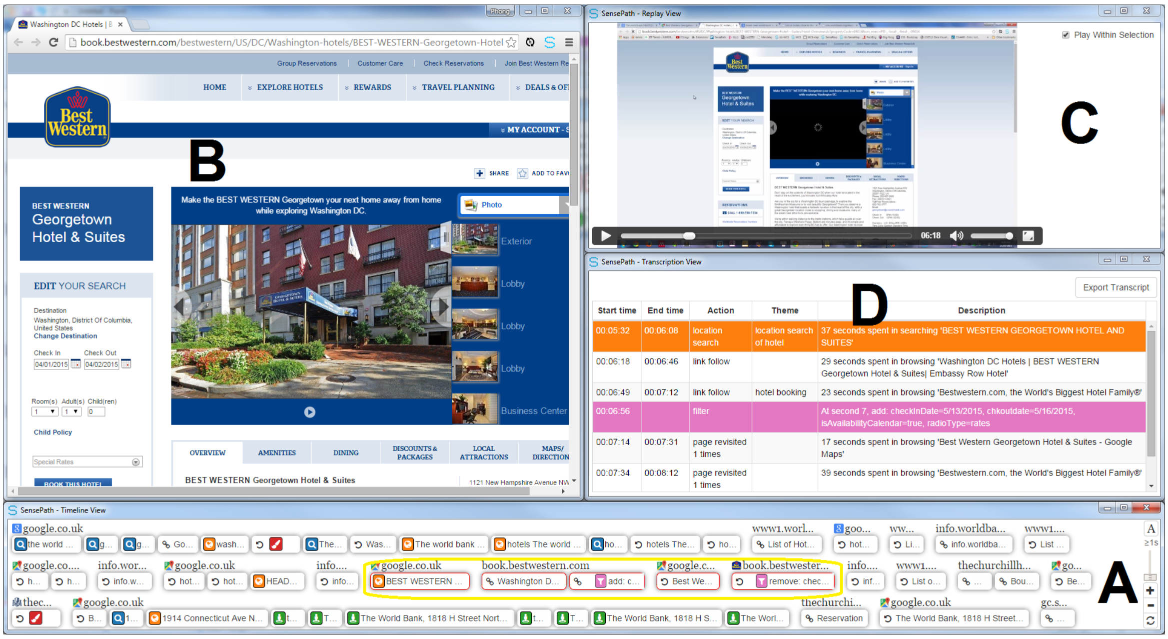


Figure 4.1: Four linked views of SensePath. **A:** The *timeline* view shows all captured sensemaking *actions* in temporal order. **B:** The *browser* view displays the web page where an action was performed. **C:** The *replay* view shows the screen capture video and can automatically jump to the starting time of an action when it is selected in another view. **D:** The *transcription* view displays detailed information of selected actions (highlighted in the timeline).

4.3.3 Provenance Capture

4.3.3.1 Content

Capture type of action, time, context, relationship. Figure 4.2 summarizes all the action types and relationships captured in SensePath.

4.3.3.2 Mechanism

Describe techniques to capture search, filter, reading actions and their limitations.

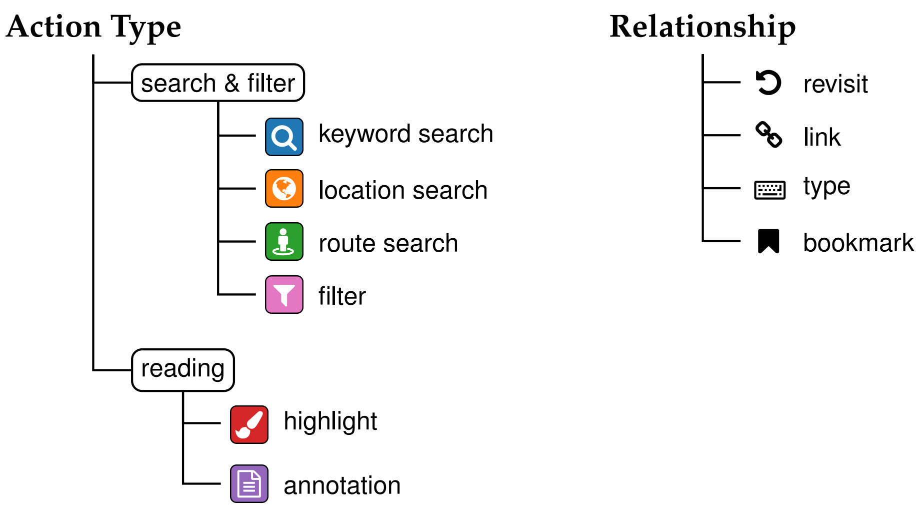


Figure 4.2: All the *action types* and *relationships* that SensePath captures, together with the icons representing them.

4.3.4 Timeline View

This view provides an overview of the entire sensemaking process showing all the captured actions in their temporal order (Figure 4.1A).

4.3.4.1 Visual Representation

Action Bar ?

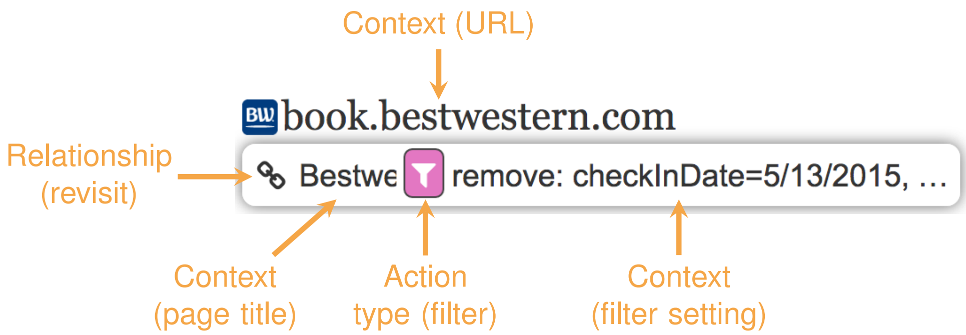


Figure 4.3: An action bar showing all four aspects of provenance information.

Action Title ?

4.3.4.2 Scalability

Different zoom levels ?

Figure 4.4: An action tile augments the action bar with a page screenshot.

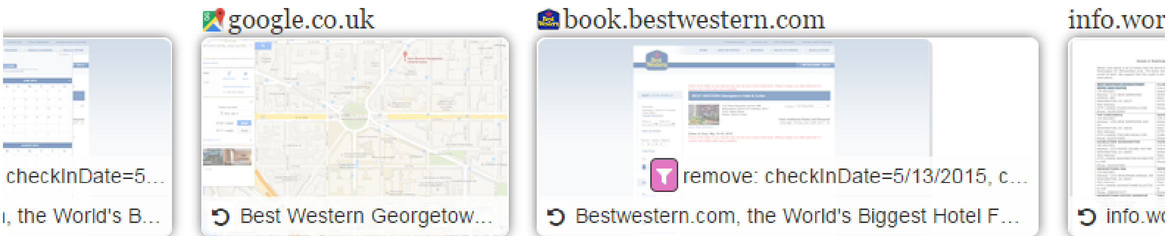


Figure 4.5: A part of the timeline with action tiles.

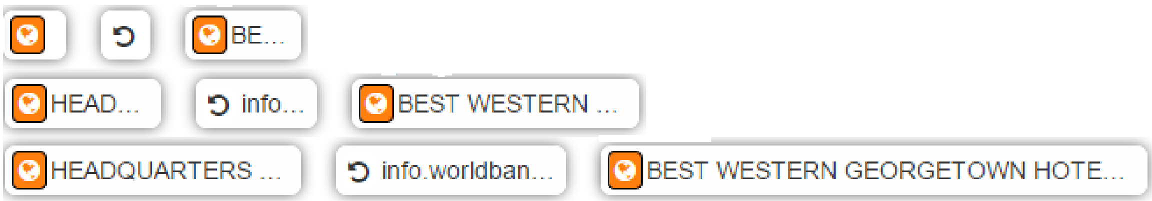


Figure 4.6: Three zoom levels of action bars with the details increasing from the top to the bottom row.

Merge action bars ?

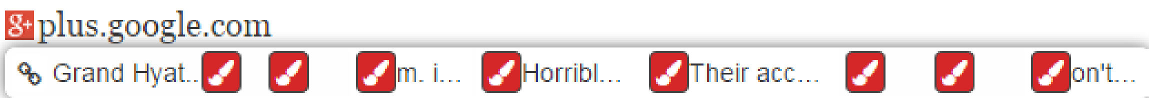


Figure 4.7: Aggregated action bar. It combines eight adjacent highlights on the same web page.

4.3.4.3 Interaction

Hovering ?

Selective zooming ?

Figure 4.8: Effect when mouse hovering an action bar (highlighted with a yellow eclipse). It highlights all other actions with the same URL (red borders) and opens a tooltip showing additional information.

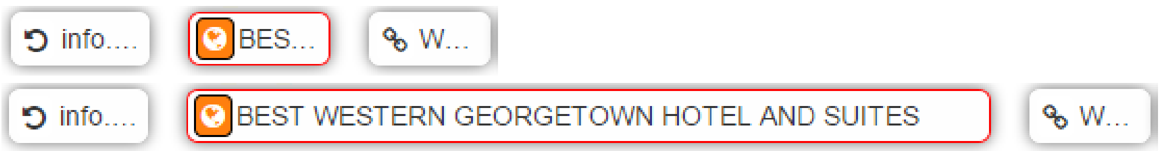


Figure 4.9: Selective zooming. Action bars with red borders are selected. Top row: action bars before zooming. Bottom row: action bars after zooming – only the selected action has its zoom level changed.

Filtering ?

Figure 4.10: Actions filtering. The slider (on the right side) controls the minimal length visible actions. Actions fall below the threshold fade out first before completely disappear.

4.3.5 Browser View

To examine the web page that a participant was looking at when performing a sensemaking action.

4.3.6 Replay View

To watch the screen capture. This view links to other views enabling the researchers to watch the portion of interest.

4.3.7 Transcription View

To provide details of selected actions. Also, support exporting the transcripts to use by many popular qualitative data analysis softwares.

4.3.8 Implementation

Discuss the Chrome extension and the two components: capture and visualization. The communication among linked views.

4.4 Evaluation

4.4.1 Evaluation 1

- Goal: Explore how SensePath is used
- Method: qualitative study
- Task: use SensePath to analyze the sensemaking process performed others
- Participants: 1, two sessions with different datasets
- Report: discuss findings, how SensePath helped

4.4.2 Evaluation 2

- Goal: Discover whether SensePath has any advantages compared to a traditional method
- Method: qualitative and quantitative
- Participants: 2
- Task: one use SensePath to analyze the sensemaking process performed others and one uses their traditional method (use a Transcribing software and Excel for coding)
- Report: discuss the difference found in the two settings in terms of time, quality of outcome of the two participants

5 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*.

5.1 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.

5.2 Design Research and Requirements

5.2.1 Design Research

We conducted a semi-structured interview with nine participants to explore how they search, manage and synthesize online information for their daily work activities.

5.2.2 Model and Design Requirements

We propose a model to describe the browser-based sensemaking process

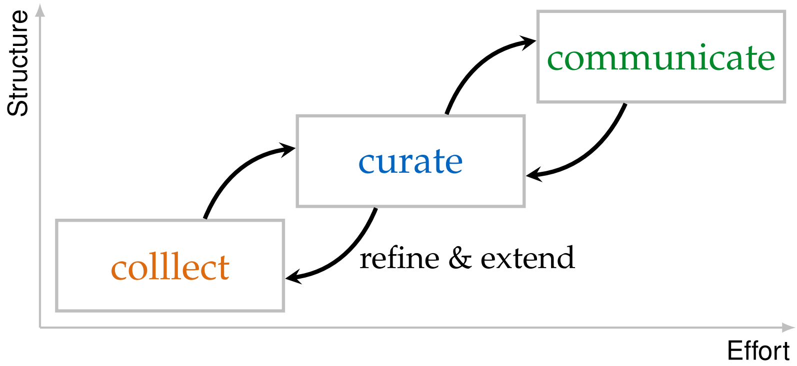


Figure 5.1: Our refined sensemaking model for user behaviors on the web.

We elicit requirements for all three processes: collection, curation, communication.

5.2.2.1 Collection Requirements

1. **Provenance visualization**: enrich and make the provenance of information sources more visible to users
2. **Easy revisitation**: ensure information sources are organized and saved; also, allow quick and easy revisitation
3. **Location awareness**: provide a good overview of the sensemaking process
4. **Preparation for curation**: enable quick relevance assessment of the information source and represent with different levels of richness depending on the relevance
5. **Interruption & Separation**: enable task switching without compromising the collection process

5.2.2.2 Curation Requirements

6. **Rich representation**: provide a rich abstraction of the information source allowing the user to quickly recognize it
7. **Free movement**: enable users to freely arrange information sources in both x and y dimensions to address the limit of a one-dimensional sequence of tabs.
8. **Causal relationship**: enable further curation of these sources by establishing links.
9. **Reasoning**: enable users to apply their reasoning strategies such as using supporting and counter evidence, or alternative hypotheses.
10. **Collection – Curation**: enable users to see connections between the curated and collected sources, and to use these to inform further searches.

5.2.2.3 Communication Requirements

11. **Complete picture**: provide a complete picture of the curated sources and the relationships that a user ascribes to them via their curation activity
12. **Raw data**: enable users to use these sources and the relationships that they have ascribed to them to communicate their findings using both the curated picture and by accessing the original sources.
13. **Varied audience**: enable users to customize the curated set of information to suit various needs and backgrounds of the audience.
14. **Share**: enable users to share both raw and curated sets of information with others.

5.3 Interface Design

5.3.1 Overview

Fig. 5.2 shows the relationship between the views and the supports they provide.

5.3.2 Browser View

This is a standard web browser with the following additional features to support sensemaking and capture provenance.

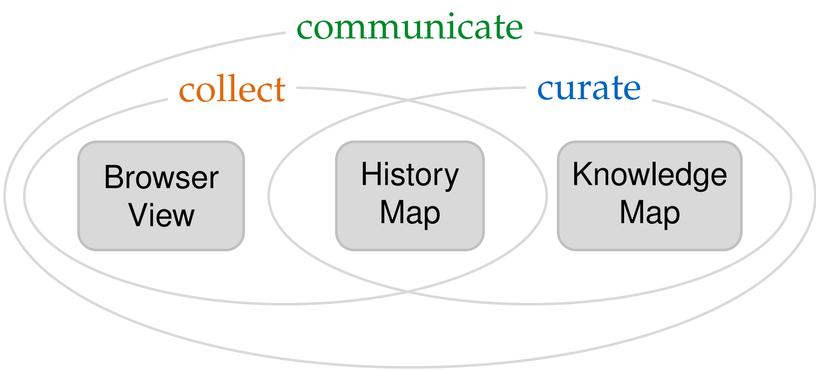
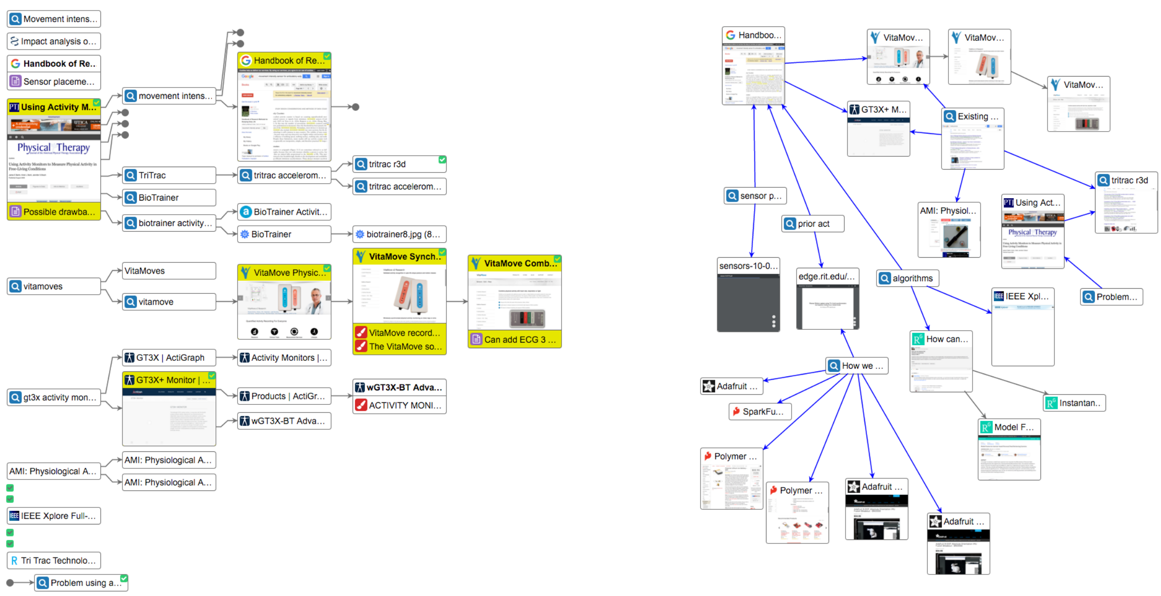


Figure 5.2: Three views and their supports.



(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 5.3: SenseMap interface.

5.3.3 History Map

This map provides an overview of the sensemaking process using the captured actions and their provenance (Fig. 5.3a).

5.3.3.1 Visual Representation

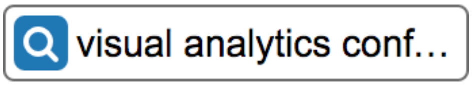


Figure 5.4: Action bar for a search with keyword "visual analytics conference".

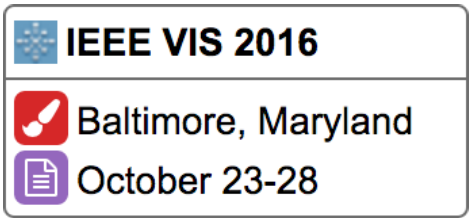


Figure 5.5: A page with one highlight and one note.

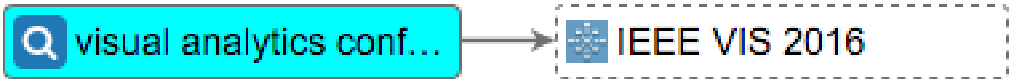


Figure 5.6: The user is active on a search result page (left bar) and opens a link in a new tab (right bar).

5.3.3.2 Layout

Discuss how the layout is generated. Explain why the branching history addresses Requirement 1 – rich provenance and Requirement 3 – location awareness

5.3.3.3 Preparation for Curation

Enable users to assess the relevance of collected information (Requirement 4). Three levels:

1. If a node is completely irrelevant, the user can *remove* it.
2. If a node is not quite relevant but the user wants to keep it to have a look at some point, they can *minimize* it.
3. If a node is very relevant, the user can *favorite* it.

Visual representation of these different levels

5.3.3.4 Scalability

Through semantic zooming – smaller level contains less information

5.3.3.5 Revisitation and Interruption

Explain how these two requirements are met

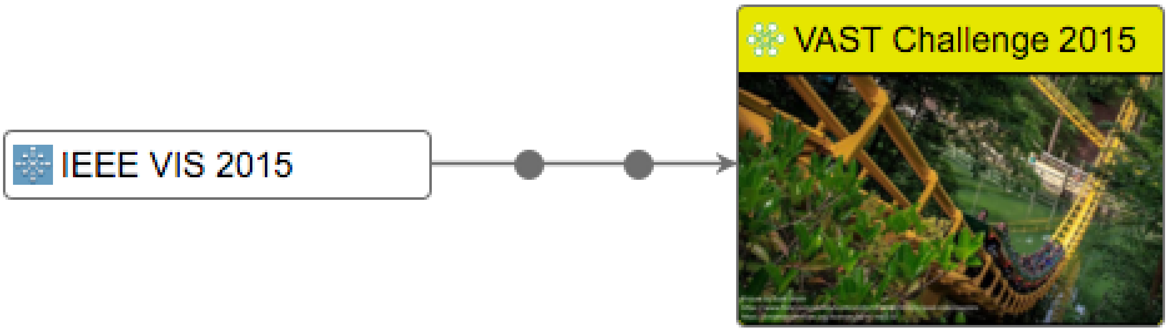


Figure 5.7: Nodes are pre-curated: two middle nodes are minimized and the last one is set favorite.

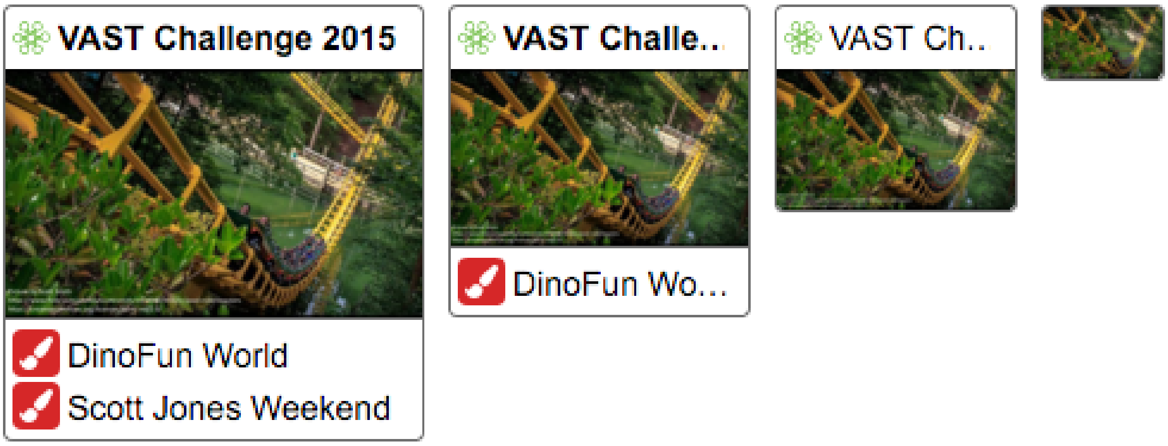


Figure 5.8: The same node with four zoom levels.

5.3.4 Knowledge Map

This map allows users to curate the information displayed in the History Map.

5.3.4.1 Visual Representation

Address Requirement 6 – rich representation

5.3.4.2 Free Movement

Address Requirement 7

5.3.4.3 Casual Relationship

Address Requirement 8

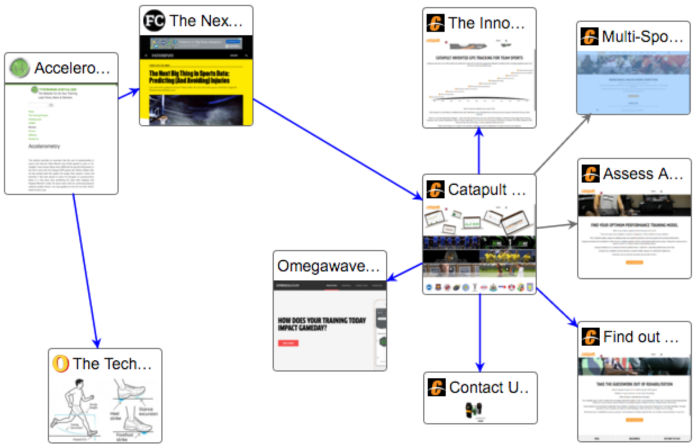


Figure 5.9: A knowledge map with three clear groups of nodes as the result of free movement.

5.3.4.4 Reasoning

Address Requirement 9

5.3.4.5 Collection – Curation

Address Requirement 10

5.3.5 Communication

Discuss how communication requirements are met.

5.4 Evaluation

- Goal: explore how SenseMap is used in a naturalistic work setting
- Method: qualitative study, backed up by some numbers
- Task: use SenseMap to perform a sensemaking task related to their daily activities
- Participants: 5
- Report: thematic analysis, discuss the common features found among particiapants, and describe a model to summarize it.