Interactive Exploration of Implicit and Explicit Relations in Faceted Datasets

Jian Zhao, Christopher Collins, Fanny Chevalier, and Ravin Balakrishnan
Jian Zhao

www.cs.toronto.edu/~jianzhao/
Faceted Document Data

- Large in Size
Faceted Document Data

- Large in Size
- Rich in Structure
  - Heterogeneous facets (attributes)
  - Numerical, categorical, and ordinal
  - Authors, Venue, Year, Keywords, …
Faceted Document Data

• Large in *Size*

• Rich in *Structure*
  ○ Heterogeneous facets (attributes)
  ○ Numerical, categorical, and ordinal

• Complex in *Semantics*
  ○ Implicit relationships
Faceted Document Data

- Large in Size
- Rich in Structure
  - Heterogeneous facets (attributes)
  - Numerical, categorical, and ordinal
- Complex in Semantics
  - Implicit relationships
  - Explicit references
Related Work - Polaris

- Visual query language
- Direct manipulation

Related Work - InfoZoom

- Filter on attribute values
- Compressed views

Related Work – Semantic Substrates

• Partitioned dataset
• Links within and across canvas areas

Related Work – Pivot Graphs

- Attribute-based organization
- Roll-up of nodes

PivotSlice

- **Data**: scholarly publications from an online database
- **Concept**: documents arranged as graph nodes in a sliceable workspace
- **Queries**: visual queries based on implicit relations (document metadata)
- **Relations**: explicit relations revealed (citations and references)
PivotSlice

Pivot

• Aggregate, sort, or align data items with any faceted attributes on the two axes
PivotSlice

Pivot
• Aggregate, sort, or align data items with any faceted attributes on the two axes

Slice
• Distribute the information space into a tabular view of sub-regions based on logical queries
PivotSlice

**Pivot**
- Aggregate, sort, or align data items with any faceted attributes on the two axes

**Slice**
- Distribute the information space into a tabular view of sub-regions based on logical queries

- Multi-scale representation
- Direct manipulation
- Smooth animation
Visual Query Logic

Data Filter

Facet panels

Query axis

Y₀

Q₀₀

Q₁₀

Q₂₀

Y₁

a₁

b₁, b₂

Q₀₁

Q₁₁

Q₂₁

X₀

b₂, b₃

X₁

a₂

m₁,₂

X₂
Visual Query Logic

\[ Y_0 \]

\[ Y_1 \]

\[ a_1, b_1, b_2 \]

\[ Q_{00} \]
\[ Q_{10} \]
\[ Q_{20} \]

\[ Q_{01} \]
\[ Q_{11} \]
\[ Q_{21} \]

\[ X_0 \]
\[ X_1 \]
\[ X_2 \]

\[ b_2 \cup b_3 \]
Visual Query Logic

\[ a_1 \cap (b_1 \cup b_2) \]

\[ (a_1 \cap (b_1 \cup b_2)) \cap (b_2 \cup b_3) \]
Information Seeking in Faceted Datasets

• Overview first
• Pivot and slice
• Relate and extract
• Details-on-demand
• History available

Information Seeking in Faceted Datasets

- *Overview first*
- Pivot and slice
- Relate and extract
- Details-on-demand
- History available
Information Seeking in Faceted Datasets

- Overview first
- *Pivot and slice*
- Relate and extract
- Details-on-demand
- History available
Information Seeking in Faceted Datasets

- Overview first
- Pivot and slice
- *Relate and extract*
- Details-on-demand
- History available
Information Seeking in Faceted Datasets

• Overview first
• Pivot and slice
• Relate and extract
• Details-on-demand
• History available
Information Seeking in Faceted Datasets

- Overview first
- Pivot and slice
- Relate and extract
- Details-on-demand
- *History available*
Information Seeking in Faceted Datasets

- Overview first
- Pivot and slice
- Relate and extract
- Details-on-demand
- History available
- Dinosaurs
Qualitative Evaluation

• Participants
  o 6 graduate students (2 females and 4 males)
  o Aged 24-33
  o Had at least 2 years of research experience
  o Had conducted literature searches as part of research
Qualitative Evaluation

• Tasks
  o 20 low-level analytical tasks, classified into 10 groups according to Amar et al. [2]
  o 3 high-level exploratory tasks

Questionnaire Results

Q1. Easy to learn.
Q2. Easy to use.
Q3. Helpful to organize and browse data.
Q4. Helpful to locate and query specific data.
Q5. Helpful to reveal and obtain data information.
Q6. Helpful to identify and interpret data relationships.
Q7. I feel the animations are useful.
Q8. I feel the graphical histories are useful.
Q9. I feel the integration of online searching is useful.
Q10. I would like to use PivotSlice to explore faceted datasets.
Observation and Feedback

• Interaction observation
  • Low-level tasks
    • All completed relatively quickly (~1.5min each)
    • Various approaches emerged for the same task
  • High-level tasks
    • Used different functions to find insights and explore more deeply in the data
Observation and Feedback

• Sample feedback
  + “[PivotSlice] makes it easy for organizing complicated filters and searches”
  + “it is extremely helpful in uncovering trends and distributions between data attributes [...] relating different parts of data is easy by using the connections and different alignments”
    – “it is sometimes confusing that you can perform the same task in many different ways”
    – “consider using similar visual cues from the field of logical circuit design [for representing AND/OR operations]”
Future Work

• Apply PivotSlice to other types of faceted data
  o Movie and music databases
  o Large scale datasets

• Conduct further evaluation
  o Laboratory and practical settings (deployment)
  o Quantitative and comparative study

• Improve the visualization and interaction techniques
  o Better node layout (currently force-directed)
  o Semantic zooming
  o Edge bundling
Interactive Exploration of Implicit and Explicit Relations in Faceted Datasets

Jian Zhao, Christopher Collins, Fanny Chevalier, and Ravin Balakrishnan

Try it yourself: www.cs.toronto.edu/~jianzhao/ or tiny.cc/pivotslice
<table>
<thead>
<tr>
<th>Tasks</th>
<th>38. What are the papers with keyword Information Visualization in conference Computer Human Interaction?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieve Value</td>
<td>T1. Who are the authors of Cone trees: Animated 3D visualization of hierarchical information?</td>
</tr>
<tr>
<td></td>
<td>T2. What are the papers cited by paper H3: Laying out large directed graph in 3D hyperbolic space?</td>
</tr>
<tr>
<td>Filter</td>
<td>T3. What are the papers coauthored by John Stasko and Robert Amar?</td>
</tr>
<tr>
<td></td>
<td>T4. What are the papers with keyword Information Visualization in 2000–2002? How do papers in these years cite each other?</td>
</tr>
<tr>
<td>Determine Range</td>
<td>T5. What are the publication years of author Catherine Plaisant?</td>
</tr>
<tr>
<td></td>
<td>T6. What is the range of citation counts for papers written by Chris North?</td>
</tr>
<tr>
<td>Find Extremum</td>
<td>T7. What are the most popular keywords of papers in conferences excluding Computer Human Interaction?</td>
</tr>
<tr>
<td></td>
<td>T8. Which author publishes the most papers in conference IEEE Symposium on Information Visualization?</td>
</tr>
<tr>
<td>Compute Derived Value</td>
<td>T9. Find all the citations of the papers with keyword Information Visualization in conference Computer Human Interaction.</td>
</tr>
<tr>
<td></td>
<td>T10. Who are the collaborators of author Stuart Card? Are there any patterns in his coauthors?</td>
</tr>
<tr>
<td>Sort</td>
<td>T11. Order papers with keywords Information Space or Dynamic Query, by year. Tell us one insight about the citation patterns among papers in different years.</td>
</tr>
<tr>
<td></td>
<td>T12. Rank papers written by author George Robertson with and without Jock Mackinlay by citation count. What is the most cited paper and how do other papers cite it?</td>
</tr>
<tr>
<td>Characterize</td>
<td>T13. What is the distribution of Stuart Card’s papers by conference? How do papers in those conferences refer to each other?</td>
</tr>
<tr>
<td>Distribution</td>
<td>T14. Compare the distributions of papers by year between conference Computer Human Interaction and IEEE Symposium on Information Visualization. What can you identify?</td>
</tr>
<tr>
<td>Find Anomalies</td>
<td>T15. Are there any exceptions for the conference IEEE Symposium on Information Visualization in terms of paper year trends?</td>
</tr>
<tr>
<td></td>
<td>T16. Observe papers with keywords Information Seeking or Dynamic Query in this dataset. Is there any incomplete data in terms of citations and references? What are the papers?</td>
</tr>
<tr>
<td>Cluster</td>
<td>T17. Identify groups of papers in conference IEEE Symposium on Information Visualization, in terms of similar year and citation count attribute values.</td>
</tr>
<tr>
<td></td>
<td>T18. Find clusters of papers written by Stuart Card, George Robertson, or Jock Mackinlay. Can you tell more about the collaboration patterns between them?</td>
</tr>
<tr>
<td>Correlate</td>
<td>T19. Compare papers containing both keywords Information Visualization and User Interface, and the ones containing neither of them. Who are the authors that publish the most in the two categories? What are the popular keywords in addition to the two above?</td>
</tr>
<tr>
<td>High Level Tasks</td>
<td>T21. Tell three facts about the author Ben Shneiderman.</td>
</tr>
<tr>
<td></td>
<td>T22. Give three comments on the conference User Interface Software and Technology.</td>
</tr>
<tr>
<td></td>
<td>T23. Name three interesting findings for the Direct Manipulation research keyword.</td>
</tr>
</tbody>
</table>
Layout - Force Directed
Layout - Aggregated
Layout - Facet-Aligned