Provenance-Based Visualization Retrieval

Holger Stitz, Samuel Gratzl, Harald Piringer, and Marc Streit

Storing interaction provenance generates a knowledge base with a large potential for recalling previous results and guiding the user in future analyses. However, search and retrieval of analysis states can become tedious without extensive creation of meta-information by the user. In this work we present an approach for an efficient retrieval of analysis states which are structured as provenance graphs of automatically recorded user interactions and visualizations. In the following, we use the Glimpse-inspired prototype from [4] as guiding example (see Fig. 1).

**Motivation**
Our collaboration partners, who are researchers at a pharmaceutical company, aim to discover cancer genes that can be targeted with future drugs. For the data-driven drug discovery, the analysts use a specialized visual analysis software that records all user interactions as provenance graphs. In such scenarios, analysts work in distributed teams and perform analyses using the same software. Hence, it is likely that an analyst wants to know if she or a colleague has already investigated the same or a similar set of genes in earlier analyses. Using our approach, the analyst can find similar visualization states by performing a retrieval based on her colleagues’ as well as her own provenance information.

**Visualizations and Interaction Design**

The user interface (see Fig. 1) consists of two views: The provenance view provides a scalable visualization of all recorded states [4] while the search view contains a search field, a list of search terms as query, a weighting editor, and a list of search results.

The search field supports the user with highlight and auto-complete of property names and values. We also support query-by-example in two ways:
- The user can add all properties of the active state (e.g., all displayed data attributes) as search terms by clicking the search button next to the state in the provenance view.
- Properties of the active state are marked in the search suggestions, where they can be spotted easily.

Added search terms can be weighted interactively by dragging the sliders in the editor.

The search results are updated automatically. When hovering over a search result, the sequence of states is highlighted in the provenance view. Selecting a search result loads the first state of the sequence.

**Visual Analysis**

Recording all user interactions results in an interaction provenance graph that contains a list of actions as edges and the corresponding visualization states as nodes (see Fig. 2a). Each visualization state is defined by properties that can be distinguished by their data type:
- **Categorical properties** include data attributes (e.g., GDP) and categorical visualization settings (e.g., axis scales).
- **Numerical properties** include derived visualization metrics, e.g., scagnostics for scatter plots.
- **Set-based properties** typically refer to selections of data items, e.g., to search for analysis states where particular countries have been brushed.

**Implementation**

The web application is based on the Phovea Platform. Demo: [vistories.org/glimpser-visualization-poster](https://vistories.org/glimpser-visualization-poster) [1].
