

Interactive Poster: HARVEST - Visualization and Analysis for the Masses

David Gotz, Zhen Wen, Jie Lu, Peter Kissa, Michelle X. Zhou
IBM T.J. Watson Research Center

Nan Cao, Wei Hong Qian, Shi Xia Liu
IBM China Research Lab

ABSTRACT

HARVEST is an intelligent visualization system designed to empower everyday business users to derive insight from large amounts of data. It combines three key technologies to support a complex, exploratory visualization-based analysis process without requiring users to be visualization or computer experts. First, HARVEST employs a set of smart visual analytic widgets that can be easily reused across applications and support incremental visual updates as required by a continuous visual analytic process. Second, it has a visualization recommendation engine that dynamically suggests suitable visualization widgets to users in context. Third, it supports the semantics-based capture of user visual analytic activity for the reuse and sharing of insight provenance.

Keywords: Automated Visualization, Visual Analytics

Index Terms: H.4.3 [Communications Apps]: Info Browsers—

1 INTRODUCTION

In recent years, a large number of visualization systems have been developed to help users view, explore, and analyze information. The capabilities supported by these visualization systems vary broadly, ranging from supporting casual visual collaborations to commercial-grade visual analytics.

At the same time, businesses are creating and storing more data than ever before. Companies are increasingly looking to visualization as a way to uncover business insights from this data in support of their employees' daily business tasks. Yet despite this requirement for visualization-based analysis from average business users, most of today's visualization systems still target two niche audiences: (1) dedicated information analysts and (2) dashboard consumers.

Dedicated information analysts are those who have already acquired a high degree of visualization and computer skills and often use sophisticated visualization software (e.g., Spotfire [2] or Tableau [1]). These tools provide great analytic power, but are typically too complex for average business users who often possess limited computer or visualization skills.

In contrast, dashboard consumers are typically casual users of visualization. They often use common visualizations to view information (e.g., business dashboards) or engage in collaboration (e.g., ManyEyes [4]). By design, these tools require far less skill and are accessible to a much wider range of users. However, they lack key capabilities, such as continuous exploration of large data sets, which are required by many real-world business tasks.

While both of these audiences benefit greatly from existing visualization tools, there is a third and perhaps largest class of users for whom existing tools are of limited value. This class of users—everyday business users—often have extensive domain knowledge but are not visualization or computer experts. Yet as part of their daily responsibilities, they perform situational information analysis tasks for which visualization can be of great benefit.

To address these challenges, we are building HARVEST, a visual analytic system that is designed for everyday business users.

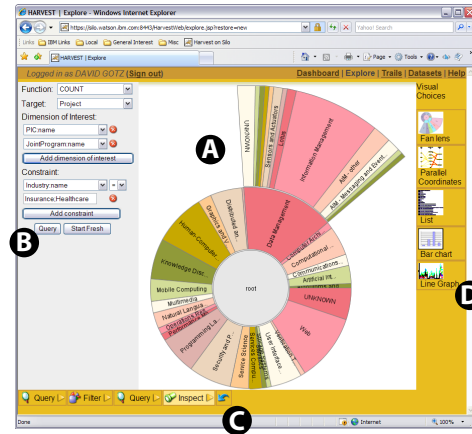


Figure 1: HARVEST is a web-based visualization system that includes (a) a visualization canvas, (b) a dynamic query panel, (c) semantic action tracking, and (d) a list of recommended visualizations.

HARVEST (Figure 1) combines three key technologies with a web-based architecture to support an exploratory visual analytic process without requiring users to be visualization or computer experts.

2 HARVEST

To help everyday business users perform complex analysis tasks, HARVEST combines three key technologies: (1) a set of smart visualization widgets, (2) dynamic visualization recommendation, and (3) semantics-based capture of visual analytic activity.

Smart Visualization Widgets: HARVEST includes a set of smart visualization widgets with standard interfaces that can be easily reused across applications. Figure 3 shows screenshots from a sampling of the visualizations used by HARVEST.

Each widget is designed to accommodate evolving data interests typical to the exploratory analysis tasks for which HARVEST is designed. As users perform further interactions with the system (e.g., a follow-up query, or drilling down to more details), widget can perform visual context management to smoothly integrate incremental data updates into an existing visualization [5].

HARVEST widgets are also designed to recognize user interactions at a semantic granularity using a standard vocabulary of user behaviors which we call *actions* (e.g., Filter, Sort, Pan, and Bookmark) [3]. This is in contrast to the common practice of using an event-based interaction model (e.g., clicks and drags). While the underlying implementations of a particular type of action may vary across visualization widgets, actions with the same type are considered semantically equivalent by HARVEST. This allows HARVEST to respond directly to a user's intent without knowing the specific inner workings of each visualization.

Dynamic Visual Recommendation: HARVEST employs a context-driven, example-based visual recommendation algorithm that assists users in finding the proper visualizations for use in their tasks. For dynamic situational tasks, it is impossible to know *a priori* which visual metaphor is most appropriate. Moreover, it can be difficult for average business users to know for themselves which choice would be best.

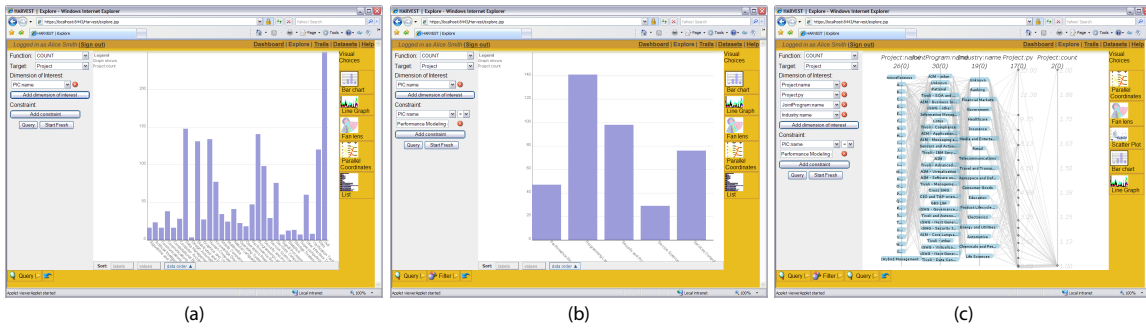


Figure 2: The typical flow of a HARVEST session. (a) The results of an initial query are displayed in a bar chart as recommended by HARVEST. (b) The user can filter down to a subset of interest by interacting directly with the visualization. (c) A follow-up query gathers additional information about the filtered subset for which HARVEST recommends a parallel coordinates visualization.

To assist in this process, HARVEST uses a visual recommendation engine to automatically recommend the top-N visualizations most suitable to a user’s ongoing task and data focus. The top recommendation is used as a default view, while alternatives are provided through the interface (see Figure 1D). A user can switch from one view to another with a single click.

Recommendations are updated dynamically throughout a user’s task based on his/her evolving data interests and behavior. For example, HARVEST actively analyzes a user’s interaction behavior in search of meaningful patterns (e.g., iteratively inspecting slices of the visualization in Figure 4) which can often signal implicit analytic needs. If such a pattern is detected, HARVEST infers the user’s analytic intent (e.g., visual comparison) and incorporates it into the visualization recommendation algorithm.

Semantic Insight Provenance: HARVEST employs a semantics-based approach to modeling and capturing a user’s analytic process. It monitors the temporal sequence of semantic analytic actions performed by a user over the course of his/her analysis and automatically builds a semantic model of the user’s logical exploration process. A segment of this model, called a Trail [3], is exposed through the user interface and represents the user’s current investigational thread.

The user interface (Figure 4) allows users to review the logical sequence of steps that have led them to their current stage of analysis. Mousing over the buttons provides small thumbnails of each previous state to provide an ‘at-a-glance’ pictorial summary of their progress. In addition, users are able to directly manipulate their trails via the interface. For example, a user can revisit previous stages of an analysis, or modify the parameters of an earlier action to re-use his/her logic within a new context.

Trails are also persisted together with traditional visualization bookmarks each time a user decides to save his/her work. The trail serves as a representation of insight provenance and is restored along with the interactive visualization when saved work is re-opened. This allows users who review a saved visualization to see not just *what* insight was found, but also *how* it was found. Moreover, restored trails can be modified to explore ‘what if’ scenarios or to pick up the analysis where the original trail left off.

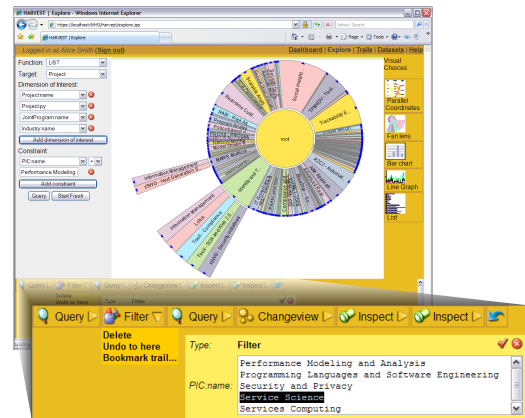


Figure 4: Editable action trails allow re-use of analytic processes in the context of new tasks. Trails are displayed on the navigation panel.

3 USE CASE

Our work on HARVEST is motivated by the common information needs of employees within our own company. We have applied HARVEST to a large collection of wiki documents describing ongoing research projects. The original wiki data allowed users to lookup information about single projects, but it was difficult to gain insights about the distribution of projects across various strategies, industries, or funding sources.

Using HARVEST, employees can easily perform ad hoc analyses over the contents of the wiki to assist in drafting their new project proposals (Figure 2). For example, to determine how best to spin a research topic, a user can quickly see which industry areas are best funded. To determine the best place to make a request for funds, the user can then analyze which funding sources are most generous within the industry identified in the first stage. To determine which executives should be lobbied for support, the user can examine which directors oversee the most projects within the target industry and funding source. All of these steps can be performed interactively using HARVEST. Moreover, the analysis can be performed via the web without special software and by users without skills in visualization, query languages, or data transformation.

REFERENCES

- [1] Tableau software. <http://www.tableausoftware.com/>.
- [2] Tibco spotfire. <http://spotfire.tibco.com/>.
- [3] D. Gotz and M. X. Zhou. Characterizing users’ visual analytic activity for insight provenance. In *IEEE VAST*, 2008.
- [4] F. B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. Many eyes: A site for visualization at internet scale. In *IEEE InfoVis*, 2007.
- [5] Z. Wen, M. X. Zhou, and V. Aggarwal. An optimization-based approach to dynamic visual context management. In *InfoVis*, 2005.

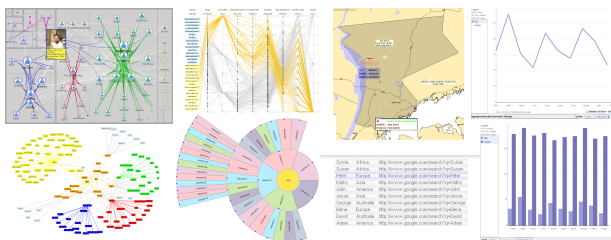


Figure 3: A sample of the visualization types included in HARVEST.