Available online at www.sciencedirect.com
ScienceDirect
Decision Support Systems 42 (2006) 1521 – 1538
Decision Support Systems
www.elsevier.com/locate/dss

Exploring auction databases through interactive visualization

Galit Shmueli [a,*], Wolfgang Jank [a], Aleks Aris [b,c], Catherine Plaisant [b], Ben Shneiderman [b,c,d]

[a] *Department of Decision and Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, USA*
[b] *Human–Computer Interaction Laboratory, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA*
[c] *Department of Computer Science, University of Maryland, College Park, MD 20742, USA*
[d] *Institute for Systems Research, University of Maryland, College Park, MD 20742, USA*

Received 11 May 2005; received in revised form 28 December 2005; accepted 18 January 2006
Available online 9 March 2006

Abstract

We introduce *AuctionExplorer*, a suite of tools for exploring databases of online auctions. The suite combines tools for collecting, processing, and interactively exploring auction attributes (e.g., seller rating), and the bid history (price evolution represented as a time series). Part of *AuctionExplorer*'s power comes from its coupling of the two information structures, thereby allowing exploration of relationships between them. Exploration can be directed by hypothesis testing or exploratory data analysis. We propose a process for visual data analysis and illustrate *AuctionExplorer*'s operations with a dataset of eBay auctions. Insights may improve seller, bidder, auction house, and other vendors' understanding of the market, thereby assisting their decision making process.

Keywords: Online auctions; Bid history; Auction dynamics; User interface; Time series

1. Introduction and motivation

Empirical research of online auctions has been flourishing in recent years due to the important role that these auctions play in the marketplace, and the availability of large amounts of high-quality bid data from websites such as eBay, Yahoo!, OnSale, and uBid. Academic researchers as well as practitioners in the eCommerce world have been trying to unveil relationships between selling, bidding and winning. In particular, a popular question has been the influence of auction design, auction dynamics, and market characteristics on the auction outcome. This information is then used to come up with recommendations for sellers, buyers, and auction sites.

In practice, there exist a variety of online services that are aimed at providing recommendations for bidders or sellers in online auctions (e.g., www.coolebaytools.com). Quite often, these recommendations are based on historical auction data and are used for strategic decision making. The online service hammertap.com, for instance, provides historical data from eBay's auctions to learn about the best-selling products, the most profitable eBay categories, and the most successful sellers and

* Corresponding author. Tel.: +1 301 405 9679.
E-mail addresses: gshmueli@rhsmith.umd.edu (G. Shmueli), wjank@rhsmith.umd.edu (W. Jank), aris@cs.umd.edu (A. Aris), plaisant@cs.umd.edu (C. Plaisant), ben@cs.umd.edu (B. Shneiderman).

0167-9236/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2006.01.001

their strategies. A similar service, Whizanalysis.com, provides complete sales analysis of the entire eBay market and identifies products and sellers that "make money." Many of these services are geared towards eBay.com, but tools for other online auction sites like Yahoo! or uBid.com also exist. Services based on historical data typically provide statistics about auctions, products, categories and sellers. These statistics arrive in the form of summary tables or simple charts. One problem is that such services do not clean the data, and therefore the historic data, and their statistics, are often based on irrelevant information (e.g., a search for a certain model of an iPod device will combine prices of both iPod devices and accessories for that model). Even if the data are manually cleaned, such summarizations are limited. While information about past transactions can be of potential strategic value to bidders and sellers, their summary in static tables and charts does not reveal their full power. Rather than summarizing the information in a static way, users would benefit from the ability to interactively explore the information. An interactive exploration would allow users to obtain an overview of all the auctions of interest and, if desired, to drill down on selected auctions and their attributes for investigating specific patterns and anomalies. Current approaches do not allow the user to explore auctions in such an interactive way.

The limited information that is provided by commercial services has led most academic researchers to collect their own raw data from online auction websites, and use these data for modeling purposes. The range of research questions that have been tested empirically is wide: multiple researchers [3,20] used eBay data to investigate the determinants of price. Ref. [4] used uBid.com data to detect types of bidding strategies. Ref. [22] compare eBay and Amazon data in terms of auction dynamics towards the auction end ("last minute bidding"). Ref. [27] used OnSale.com and SurplusAuction.com data to model the arrival times of bidders to online auctions. Ref. [10] examined eBay's reputation mechanism to learn about trust between buyers and sellers. Refs. [3] and [7] used eBay data to compare bidding strategies in auctions for common versus private value items. Ref. [16] used eBay data to study the price evolution and its dynamics, showing that dynamics differ markedly even within comparable auctions for the same good. Finally, Ref. [18] use eBay data to detect the fraudulent act of bid shilling (a seller placing dummy bids on their auction to raise the price). These are just a few of the many empirical studies that have appeared in the online auction literature.

Empirical research in online auctions has focused almost exclusively on collecting data and/or fitting statistical or other analytical models to the data, while exploratory data analysis (EDA) has been rarely addressed. EDA is a crucial first step in any data analysis and can aid in detecting data-anomalies or interesting patterns in the data. Typical tools for EDA are of static nature (summary statistics, tables and charts) and require the user to specify in advance which relationships are to be scrutinized (e.g., the user has to specify to the software first that a table of X vs. Y is desired which is then produced). This approach requires the user to have some advance knowledge of interesting patterns and will therefore not be as powerful in detecting surprising patterns. An interactive approach can be more powerful and can lead to a better understanding of the database. This is especially so for auction data which are typically of complicated nature. Current static tools will not fully display the information in auction databases for the following reasons:

(1) the need to accommodate two forms of data: typical auction data include *auction attributes* (such as seller's design choices, item attributes, seller information) and the *bid history* that describes the bids as they arrive during the auction. Users often need to study relationships across these differing data forms. In classical time series analysis, there is usually a single or a small number of time series that are of interest. In contrast, here we have a large number of auction attributes that are coupled with bid histories. Existing statistical summaries and graphical methods tend to cater either to cross-sectional data for showing auction attributes (histograms, summary statistics), or to time series (such as time plots) for showing bid histories. Combined visualizations are not standard. Our goal is to contribute in this aspect.
(2) the unique structure of bid histories makes it difficult to apply existing time series analysis tools. The bid history is a non-standard time series: it consists of bids placed at unevenly spaced time points with sparse and dense time periods. Furthermore, if we look at many auctions simultaneously, these time series differ markedly: auctions have different durations, bids in different auctions are placed at unrelated time points, and the total number of bids can vary vastly.

In a recent paper, Ref. [23] introduced several general graphical techniques for visualizing online auction data

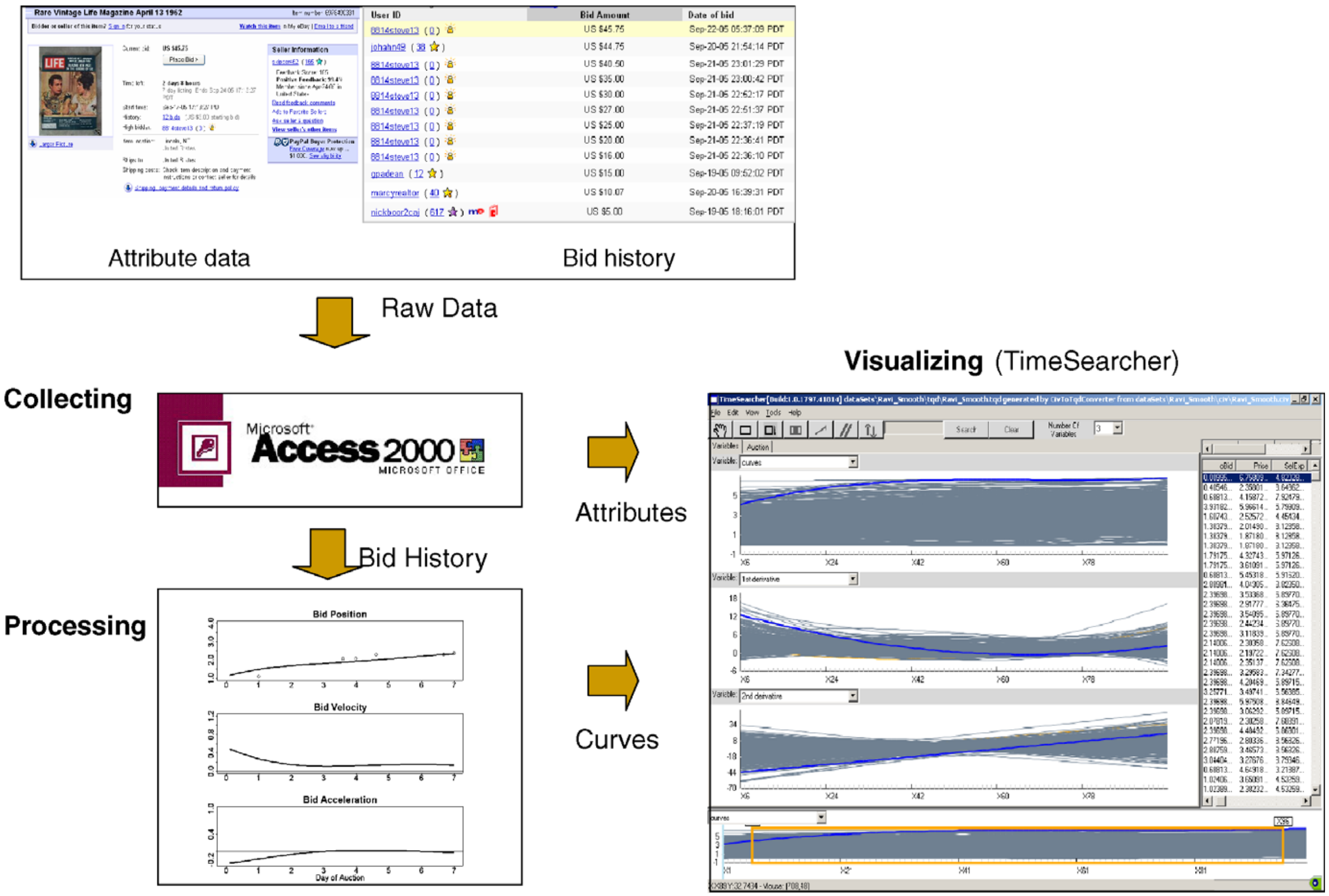


Fig. 1. The *AuctionExplorer* suite of tools. Raw auction data are collected and stored in a database. The time series information is processed into price curves. Finally, the linked time series and attribute data are explored in an interactive visualization tool.

from different views. In particular, they describe and illustrate the special characteristics of bid histories, and how they can be portrayed graphically. In one type of plot called the “profile plot”, they combine attribute information with bid history information by representing auctions as step functions and using color/symbols to integrate attribute information. They also introduce the concept of STAT-zoom, related to obtaining on-the-fly data summaries (e.g., boxplots) of prices at various relevant time scales. In contrast, this paper focuses on using interactive visualization for the purpose of decision making. We expand beyond the raw data and use new data representations to capture price evolution and its dynamics throughout an auction. We also heavily rely on interactivity, and in this context present an actual tool. Finally, all these are done in the context of large databases of auctions.

In this paper we describe a suite of tools that collect auction data, pre-process them and support interactive visualization and exploration, with a special focus on the visualization tool. The growth of online marketplaces has introduced a surging amount of publicly available data on the Internet. Although the data are now more accessible than before, using them for decision making requires collecting, organizing, and exploring them. The *AuctionExplorer* suite of tools consists of these three components, as depicted in Fig. 1. Starting from source report documents (here HTML pages from the eBay website), the data are transferred into a database (here Microsoft Access). The time series data require a processing step that leads to price curves. These curves, together with the auction attributes, are then imported into the visualization tool for interactive exploration. In addition to traditional time series data, this system integrates cross-sectional data that are tightly coupled with the time series.

We use a sample dataset of online auctions to show the different features of *AuctionExplorer*. The dataset consists of 34 magazine auctions on eBay that took place during the fall of 2004. The data include the bid histories and the attributes for each auction. The auctions start at different times and have different durations. *AuctionExplorer* aligns their time scales so that the *x*-axis shows the proportion of the auction duration. The attributes include day and time of auction closing, the auction duration, the opening and winning bids, the number of bids, the seller and buyer usernames and ratings, shipping cost, and the description given by the seller. It is easier to convey and illustrate the visualization tool using a relatively small dataset in this static paper environment. However, the suite is designed to handle larger datasets (see Section 4).

The paper is organized as follows: Section 2 describes the process of data collection and pre-processing. Section 3 describes the visualization tool and proposes an exploration process that provides exploration guidelines for data familiarization and more hypothesis-directed investigation. Section 4 analyzes the performance of TimeSearcher, the visualization component of *AuctionExplorer* suite, and Section 5 concludes the paper.

2. Data collection and processing

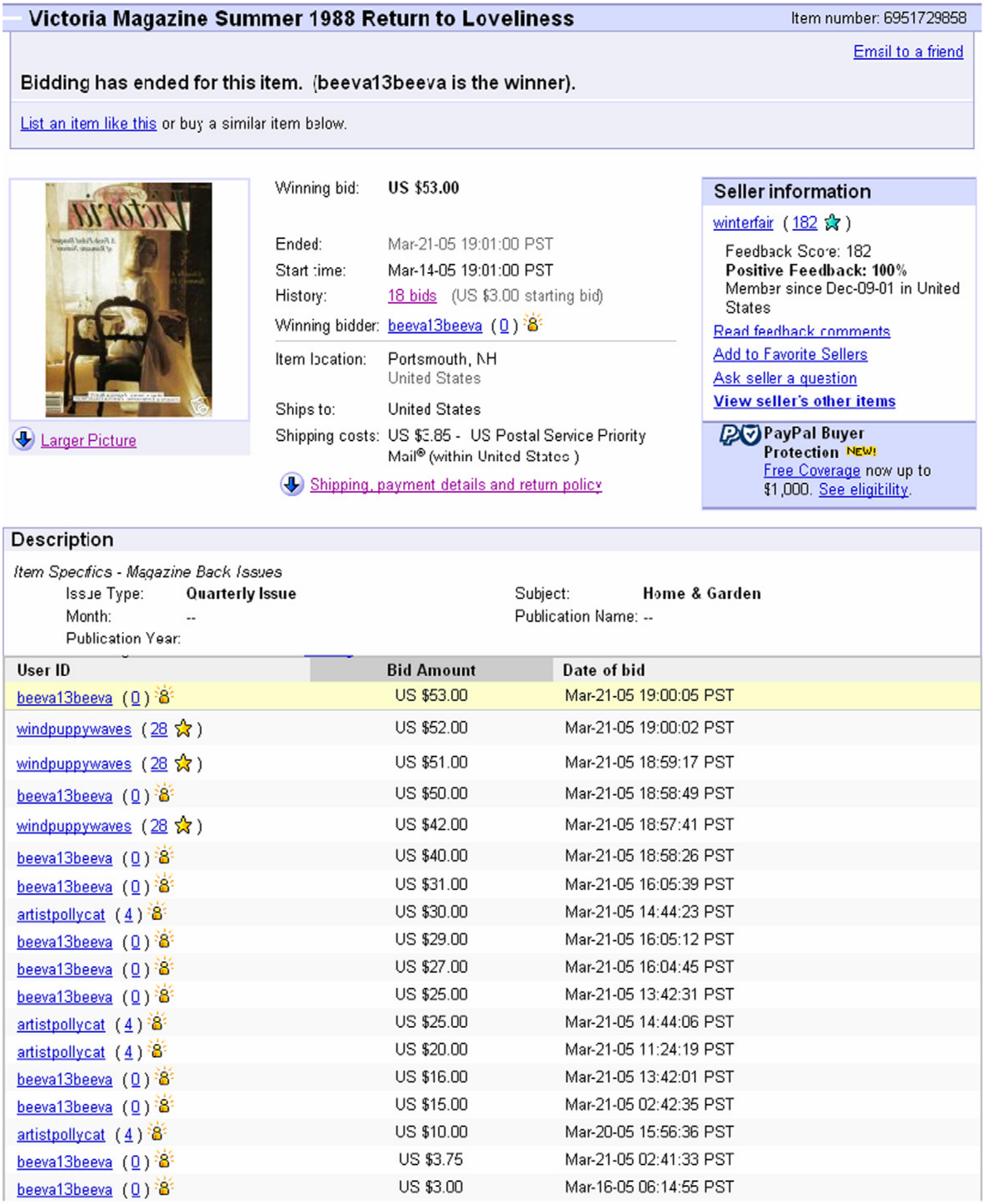
Data collection, data organization, and pre-processing steps are automatable, but require care to ensure high quality and useful auction attributes and bid histories.

2.1. Web-spiders: collecting auction attributes and bid histories

eBay.com is the largest C2C (Consumer-to-Consumer) online auction site offering millions of items for sale across thousands of product categories. Since eBay archives detailed records of its auctions completed in the last 30 days, it is a source of immense amounts of high quality data. For each closed auction the publicly available data include two parts: the first part contains the auction attributes. It includes item-specific information (category, name, and optional text and photos), auction-design parameters set by the seller (opening price, auction duration), information about the seller and winner (their rating which corresponds to their number of eBay transactions), and the closing price. We call these auction *attribute data*. The second part is a detailed log of the bidding history. For each bid the data include the bid amount, the time stamp, and the bidder username with his/her rating. We call this *time series data.* The link between an auction’s attributes and time series is the item number, a unique number that is given to each auction. Fig. 2 shows snapshots of the two HTML pages that contain the attribute data (top) and time series data (bottom) for a magazine auction. The same structure exists in other online auctions such as Yahoo! and uBid.com.

Bid amounts, called “proxy bids,” are not necessarily monotonically increasing. This is due to the second-price format of these auctions, where the second highest bid is disclosed at any point in time, but not the highest bid. We will return to this point in Section 2.2.

The use of web agents (“web spiders”) facilitates the creation of large databases of bidding data in short time. A web agent is a software application, typically written


Victoria Magazine Summer 1988 Return to Loveliness

Item number: 6951729858

Email to a friend

Bidding has ended for this item. (beeva13beeva is the winner).

List an item like this or buy a similar item below.

Larger Picture

Winning bid: US $53.00

Ended: Mar-21-05 19:01:00 PST

Start time: Mar-14-05 19:01:00 PST

History: 18 bids (US $3.00 starting bid)

Winning bidder: beeva13beeva (0)

Item location: Portsmouth, NH United States

Ships to: United States

Shipping costs: US $3.85 - US Postal Service Priority Mail® (within United States)

Shipping, payment details and return policy

Seller information

winterfair (182)

Feedback Score: 182

Positive Feedback: 100%

Member since Dec-09-01 in United States

Read feedback comments

Add to Favorite Sellers

Ask seller a question

View seller's other items

PayPal Buyer Protection NEW!

Free Coverage now up to $1,000. See eligibility.

Description

Item Specifics - Magazine Back Issues

Issue Type: Quarterly Issue

Subject: Home & Garden

Month: --

Publication Name: --

Publication Year:

User ID	Bid Amount	Date of bid
beeva13beeva (0)	US $53.00	Mar-21-05 19:00:05 PST
windpuppywaves (28)	US $52.00	Mar-21-05 19:00:02 PST
windpuppywaves (28)	US $51.00	Mar-21-05 18:59:17 PST
beeva13beeva (0)	US $50.00	Mar-21-05 18:58:49 PST
windpuppywaves (28)	US $42.00	Mar-21-05 18:57:41 PST
beeva13beeva (0)	US $40.00	Mar-21-05 18:58:26 PST
beeva13beeva (0)	US $31.00	Mar-21-05 16:05:39 PST
artistpollycat (4)	US $30.00	Mar-21-05 14:44:23 PST
beeva13beeva (0)	US $29.00	Mar-21-05 16:05:12 PST
beeva13beeva (0)	US $27.00	Mar-21-05 16:04:45 PST
beeva13beeva (0)	US $25.00	Mar-21-05 13:42:31 PST
artistpollycat (4)	US $25.00	Mar-21-05 14:44:06 PST
artistpollycat (4)	US $20.00	Mar-21-05 11:24:19 PST
beeva13beeva (0)	US $16.00	Mar-21-05 13:42:01 PST
beeva13beeva (0)	US $15.00	Mar-21-05 02:42:35 PST
artistpollycat (4)	US $10.00	Mar-20-05 15:56:36 PST
beeva13beeva (0)	US $3.75	Mar-21-05 02:41:33 PST
beeva13beeva (0)	US $3.00	Mar-16-05 06:14:55 PST

Fig. 2. Data for a closed 7-day auction on eBay for a vintage magazine: the top includes attribute data (auction attributes), and bottom includes the time series of bids (bid history).

in a programming language like Perl or Java that "crawls" over a collection of web pages and gathers the desired information. Data on thousands of auctions can be collected in minutes. This modern automated collection system is much less prone to human error than traditional data collection and recording. Advanced web agents developed by researchers can be used for the data collection step [1,6,15,19]. For our analysis, we use a web agent that uses PHP and MySql [6]. The resulting data are organized in a database.

2.2. Pre-processing the time series component: from bids to price curves

The bidding history, which reflects the price increase throughout the auction, is a time series, but it deviates from traditional time series in four main aspects: the duration of the time series is limited to a closed interval, the time points are unevenly spaced, the number of bids varies across auctions, and different time series have bids at different time points. The last three pose a challenge for most visualization and analysis tools, which expect evenly spaced series [2]. One of our solutions, which we use in *AuctionExplorer*, is to move from unequally spaced points to a continuous curve, which can then be sampled at evenly spaced intervals. There are other solutions to represent unevenly spaced time series in evenly spaced form. We explored four of these: event index, interleaved event index, sampled event, and aggregated sampled event [2]. There are multiple possible transformations ranging from crude interpolation within each auction to more sophisticated smoothing. In all cases, the discrete data are transformed into "price curves". The underlying assumption is that there is a family of curves that can represent a flexible population of price increases. This is directly related to the concept of a functional object in the field of functional data analysis [21], where a smoothing step based on some basis functions is used.

The pre-processing component in *AuctionExplorer* consists of two steps: first, the raw "proxy bids" are converted into "current prices", which represent the price shown by eBay at any point during the auction. As pointed out above, the proxy bids are not necessarily monotonically increasing in time. Using proxy bids poses two problems: first, from an economic point of view, this series does not describe the actual information that bidders are using during an ongoing auction. The "current price" that eBay displays is in fact the second highest bid (plus an increment) and it is the current price which bidders use for their decision making regarding their next move. Secondly, the series of proxy bids can have sharp drops in values,[1] which are hard to capture by a smooth curve. For these two reasons, we use the series of "current prices" posted by eBay during the live auction. This series, a step function, forms a monotone series and reflects the price climbing throughout the auction. Computing the current price from the proxy bid information is possible using eBay's increment table (pages.ebay.com/help/basics/g-bid-increment.html, last retrieved March 2005). An example is given in Fig. 3.

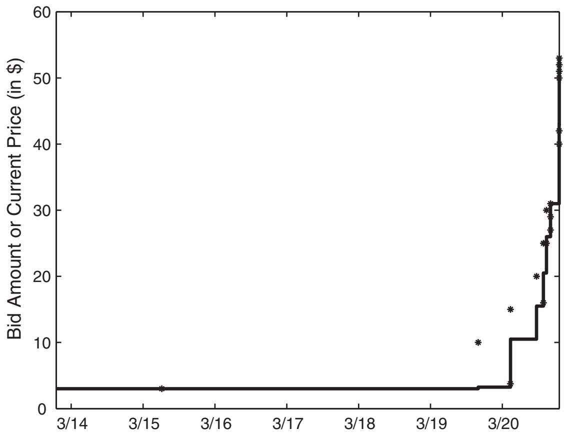


Fig. 3. Proxy bid (stars) and the current price step function for the 7-day magazine auction in Fig. 2.

Sampling the step functions at even intervals produces "well-behaved" time series that represent price evolution at regular intervals [2]. Then we represent the step function by a smoother curve to facilitate exploration of price increases and also auction dynamics. We use penalized smoothing splines[2] since they achieve good balance between data fit and excessive oscillations, and because they are computationally cheap (for further details on fitting penalized smoothing splines to auction data, related issues and alternatives, see Refs. [16] and [17]). This yields a very flexible family of curves that have the powerful feature of allowing the estimation of curve derivatives, which are of particular interest in the context of price formation [24]. The first derivative curve corresponds to the velocity of price increase. The second derivative captures the acceleration in price. The price curve and its two derivatives convey information about the price dynamics during an auction. Fig. 4 shows a bidding history with its estimated price curve (top) and its two derivative curves (bottom). It can be seen that the price in this auction goes up especially towards the auction end, and this increase has a peak acceleration around day 6.5, with a deceleration as the auction comes to a close.

[1] Bidders during the live auction only see the *second* highest bid (plus increment). Thus, it often happens that they submit a bid which is much lower than the highest proxy bid. Graphically, this leads to "drops" in the time series.

[2] We fit curves to log(price) rather than price. This achieves better fit due to the surge of activity at the last moment of the auctions, and it also better separates between curves for enhancing visualization.

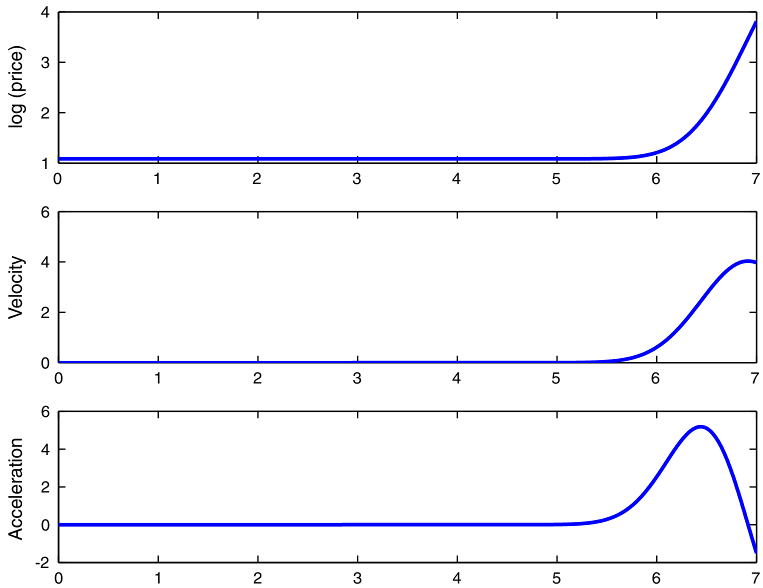


Fig. 4. Estimated price curve (in log-scale) and its first two derivatives for the magazine auction in Figs. 2, 3.

The final product of the pre-processing step is a set of three curves per auction, which are sampled on an evenly spaced grid. For our magazine example this corresponds to $3 \times 34 = 102$ curves, each sampled to 100 time points for a total of 10,200 time series data points. Notice that all auctions are aligned on a time scale that represents the proportion of the auction duration. If the actual auction duration is important, users can use the auction duration attribute to choose auctions of certain duration only (there is a limited number of possible durations on eBay: 1, 3, 5, 7, and 10 days). Next we will describe the capabilities of the interactive visualization tool, and then propose a process for exploring a dataset using these capabilities.

3. Interactive visualization and exploration

The visualization component of *AuctionExplorer* is based on TimeSearcher, a time series visualization tool developed at the Human–Computer Interaction Laboratory of the University of Maryland [8,14,13]. TimeSearcher enables users to see an overview of long time series (>10,000 points), view multivariate time series, select with rectangular time boxes, and search for a selected pattern. TimeSearcher was extended for *AuctionExplorer* to include attribute data-browsing with tabular views and filtering by attribute values and ranges (e.g., starting date or seller), both tightly coupled to the time series visualization. The application, with sample datasets, is available for download from http://www.cs.umd.edu/hcil/timesearcher. Fig. 5 shows the main screen of the visualization tool with our example dataset of 34 eBay auctions for magazines.[3] The time series are displayed in the left panel, with 3 series (i.e. 3 variables) for each auction: “Price” (top), “Velocity” (middle), and “Acceleration” (bottom), which correspond to the price curves and their first and second derivatives, as explained in the previous section. At the bottom of the screen, an overview of the entire time period covered by the auctions is provided to allow users to specify time periods of interest to be displayed in more detail on the left panel. On the right, the attribute panel shows a table of auction attributes. Each row corresponds to an auction, and each column to an attribute, starting with the auction ID number. In this dataset there are 21 attributes, scrolling provides access to attributes that do not fit into the available space. Users can choose how much screen space is allocated for the different panels by dragging the separators between the panels, enlarging some panels and reducing others. All three panels are tightly coupled so that an action in one of the panels is immediately reflected in the other panels. Attributes are matched with time series using the auction ID number as a link.

3.1. Interactive operations

The interactive visualization operations can be divided into time series operations (bid history) and

[3] A short video that illustrates the different components and possible operations of the visualization tool is available at http://www.cs.umd.edu/hcil/timesearcher/videos/ts2_HCILsoh2005R.html.

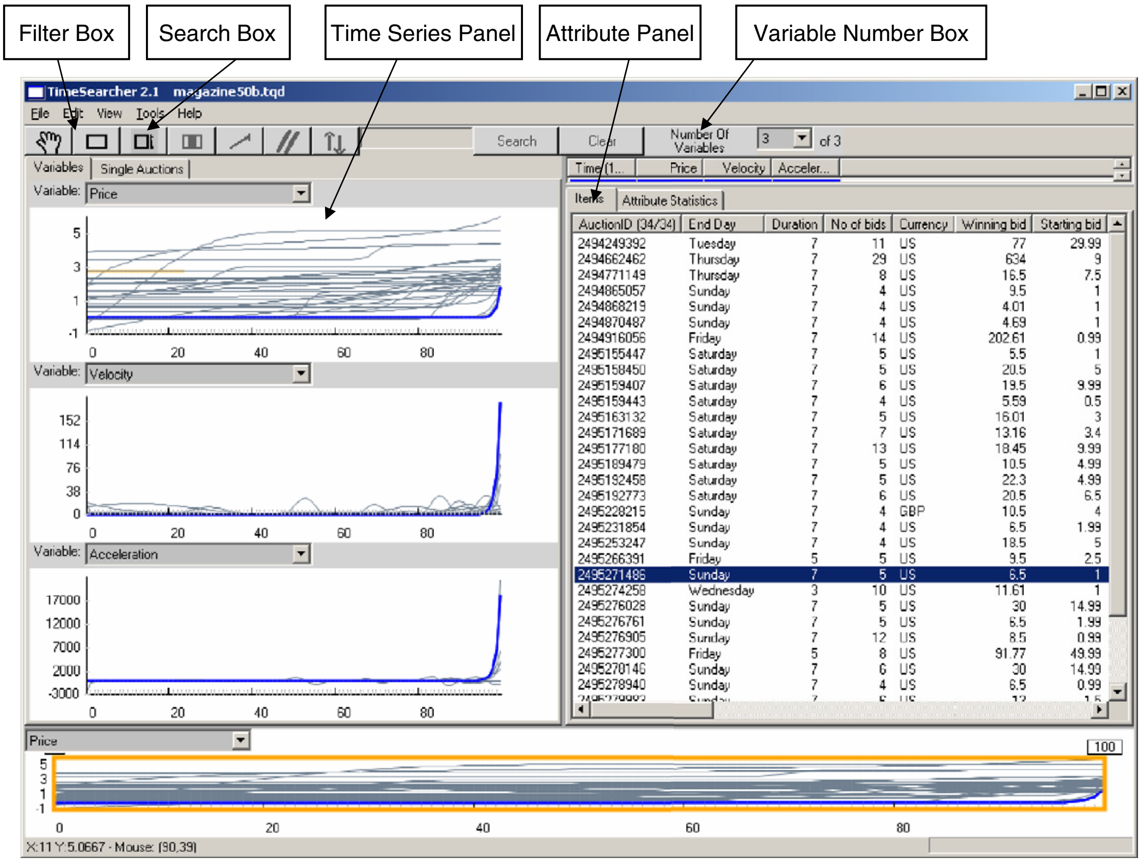


Fig. 5. Main screen of the visualization tool loaded with the sample dataset of 34 magazine auctions.

attribute operations (auction attributes). We describe these and show how they can be applied to reveal the structures and behaviors in the sample dataset.

3.1.1. Time series operations

- Zooming: The overview panel at the bottom of the screen displays the time series for one of the variables and allows users to specify in which part of the time series they want to zoom in. The orange field of view box determines the time range that is displayed on the 3 upper left panels. To zoom, users drag the sides of the box. By zooming in, the user can focus on a specific period in the data and see more details. In many cases zooming also results in better separation between the curves, enabling easier selection and un-selection of lines. This panel can display any one of the 3 time series panels. For example, Fig. 6 shows a zoomed view of the end of the auctions making clearer the difference in the end-of-auction velocity across auctions. The box can also be dragged right and left to pan the display and show a different time period. Regardless of the range of the detail view, the overview always displays the entire time series and provides context for the detail view.
- Focusing on a variable: To focus on a certain variable (price, velocity, or acceleration curves), users can choose to view only that panel on the left, thereby getting a larger view of those curves. This results in clearer separation between curves, which can be especially useful when there are many auctions. Users can specify the number of variables to be shown (here 1, 2 or 3) and select which of the variables should be displayed. Fig. 6 shows all 3 variables, while Fig. 7 shows only the price curves.
- Filtering curves: Users can filter the curves to see only auctions of interest by using filter widgets called TimeBoxes. One can click on the TimeBox icon of the toolbar and draw a box on the time series panel of interest. Every curve that passes through the box (between the bottom and top edges of the box for the duration that the box occupies) is kept while all the other curves are filtered by greying-out. The corresponding auctions are also removed from the attribute panel on the right. Fig. 8 shows a typical

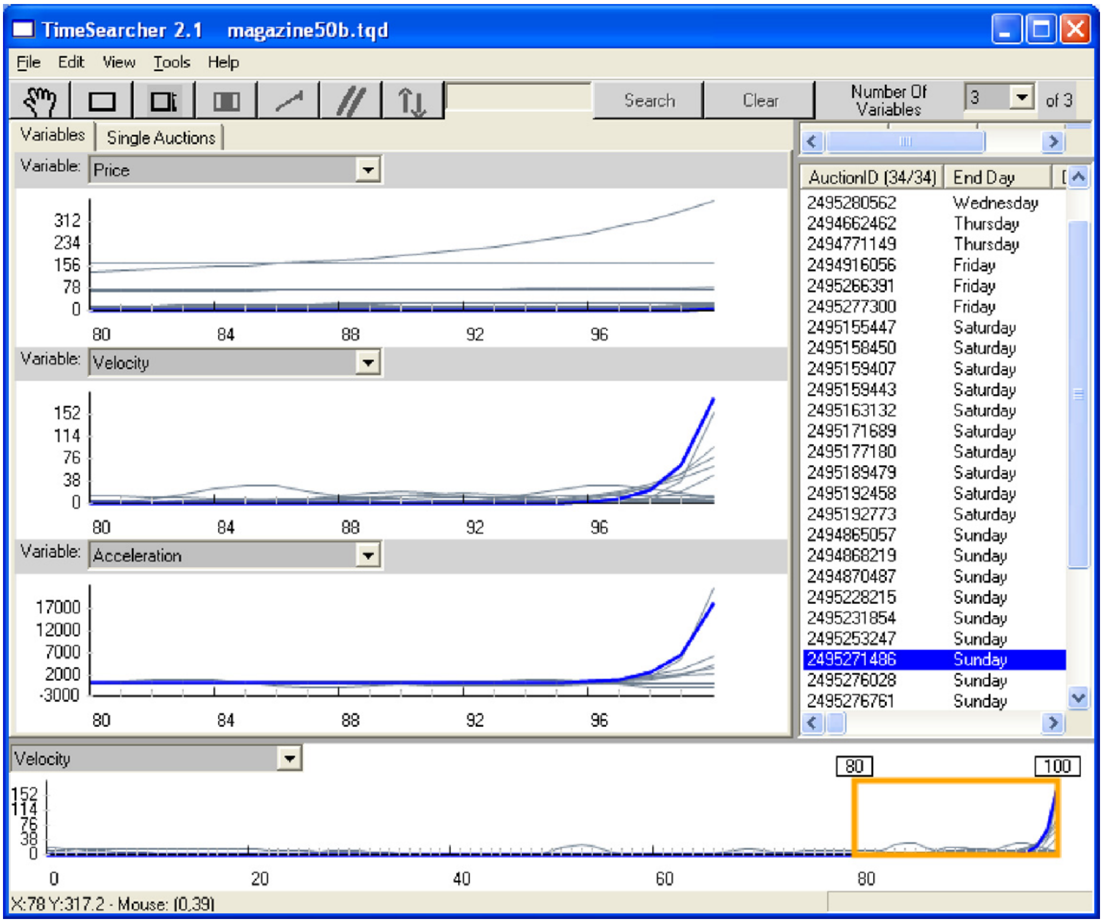


Fig. 6. Zooming in to the end of auctions to see the changes in velocity. The orange field of view displayed on the right end side of the overview allows users to specify the time period to be shown in more detail.

filter TimeBox used to see only auctions that end with high price velocity. In the attributes panel users see that they all ended around the weekend. They can apply multiple TimeBoxes on the same or separate variables, which form conjunctive queries (i.e. a combination of the query of individual TimeBoxes via logical AND). For example, users could search for auctions ending with low prices and with high velocities.

- Searching for patterns in curves: In comparing price curves, and even more so, price dynamics, a useful tool is the pattern search. This is achieved by drawing a SearchBox on a selected curve during a certain time duration. The pattern is the part of the series that the SearchBox horizontally covers, and it is searched across all other curves not only at the same time but also at any time point in the auction. There is a tolerance handle on the right of the SearchBox that allows setting a measure of similarity. For example, users can search for auctions that have price curves with steep escalations at any time during the auction (Fig. 9). The pattern search finds several auctions with the same pattern of steep increase. Strikingly, price increases occur at different auction phases: most occur at the end, but there is one that occurs at the auction start, and another at mid-auction. TimeBoxes and SearchBoxes can be combined into a multi-step interactive search [8].

3.1.2. Attribute operations

- Sorting auctions: Users can sort the auctions by any attribute by clicking on the attribute name in the 1st row. A click sorts in ascending order, while the next click sorts in descending order. Sorting can be performed on numerical as well as text attributes. The sorting also recognizes day-of-week and time formats. The sorting is useful for learning about the ranges of the values for the different attributes, the

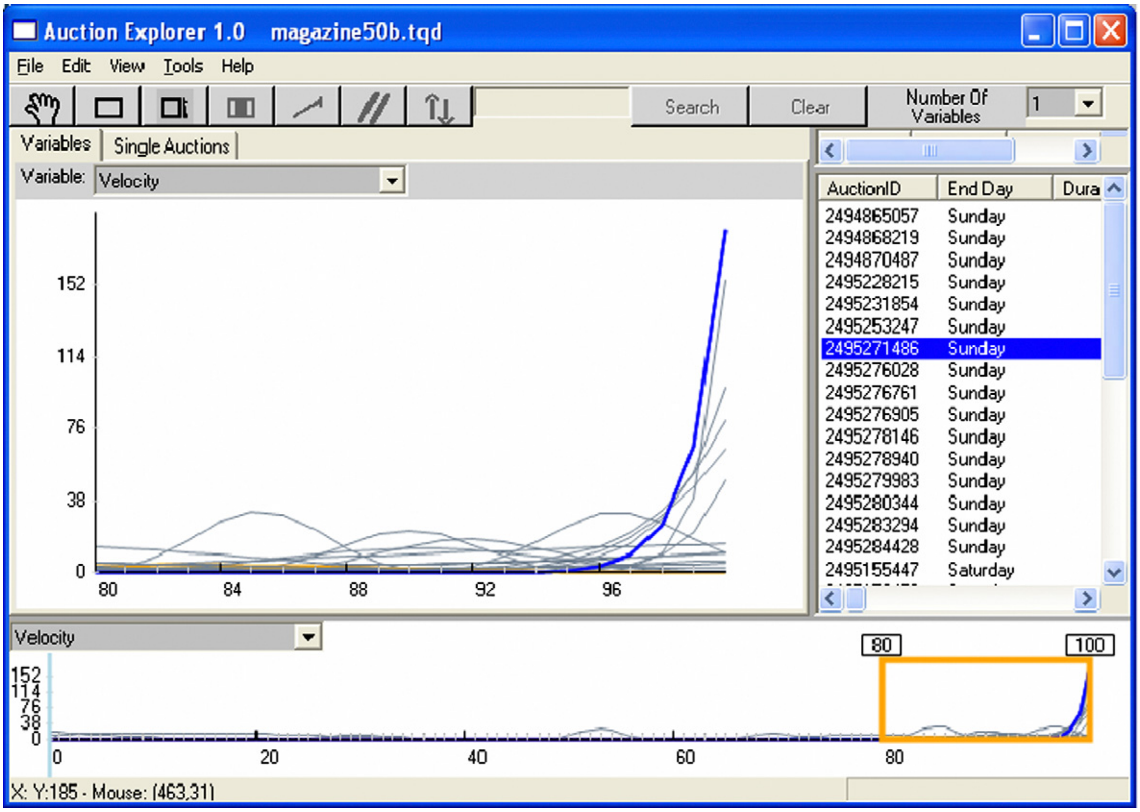


Fig. 7. Only the price curves are displayed, for a better curve separation.

existence of outliers, the absence of certain values, and possible errors and duplications in the data. Furthermore, sorting might allow users to visually spot patterns of "similar" auctions, by making auctions with similar values for an attribute consecutive in the auction list. Users may sort according to more than one column. Specifically, they can sort values first by a column and second by another column. This effect is achieved by users' clicking the second column header first, which will sort the values according to the values on the second column, and then the first column header, which will sort all attributes again according to the values in the first column. The result will be values sorted first by the first column and then the second column. Users may achieve this type of sorting for more than two columns. In addition, the order of the attribute columns can be changed by clicking and dragging the attribute names to the right or left.

- Highlighting groups of auctions: After the attribute/s of interest have been sorted, groups of auctions can be selected and their corresponding time series in the left panels are highlighted. For example, if the attributes' table is sorted by the end day of auctions, it is easy to select all auctions that end on a weekday from the table, and see the corresponding time series highlighted, revealing that they mostly comprise auctions that end with the highest prices (Fig. 10).
- Summary statistics: The summary statistics tab shows mean, standard deviation, minimum, maximum, median, and quartiles for each attribute for the selected auctions. This is updated interactively when the auctions are filtered with TimeBoxes, or when users select a subset of auctions manually. For example, while the median seller rating of all auctions is 615, when users apply a TimeBox to select auctions that started with a low price, the median seller rating jumps to 1487. Moving the TimeBox to select auctions that started with a high price results in a median seller rating of 243, which provides evidence for the fact that setting auctions with a low starting price is a strategy mostly employed by experienced sellers.

3.2. A process for visual data analysis of auction data

The interactive operations of TimeSearcher allow users to freely explore a dataset but our experience

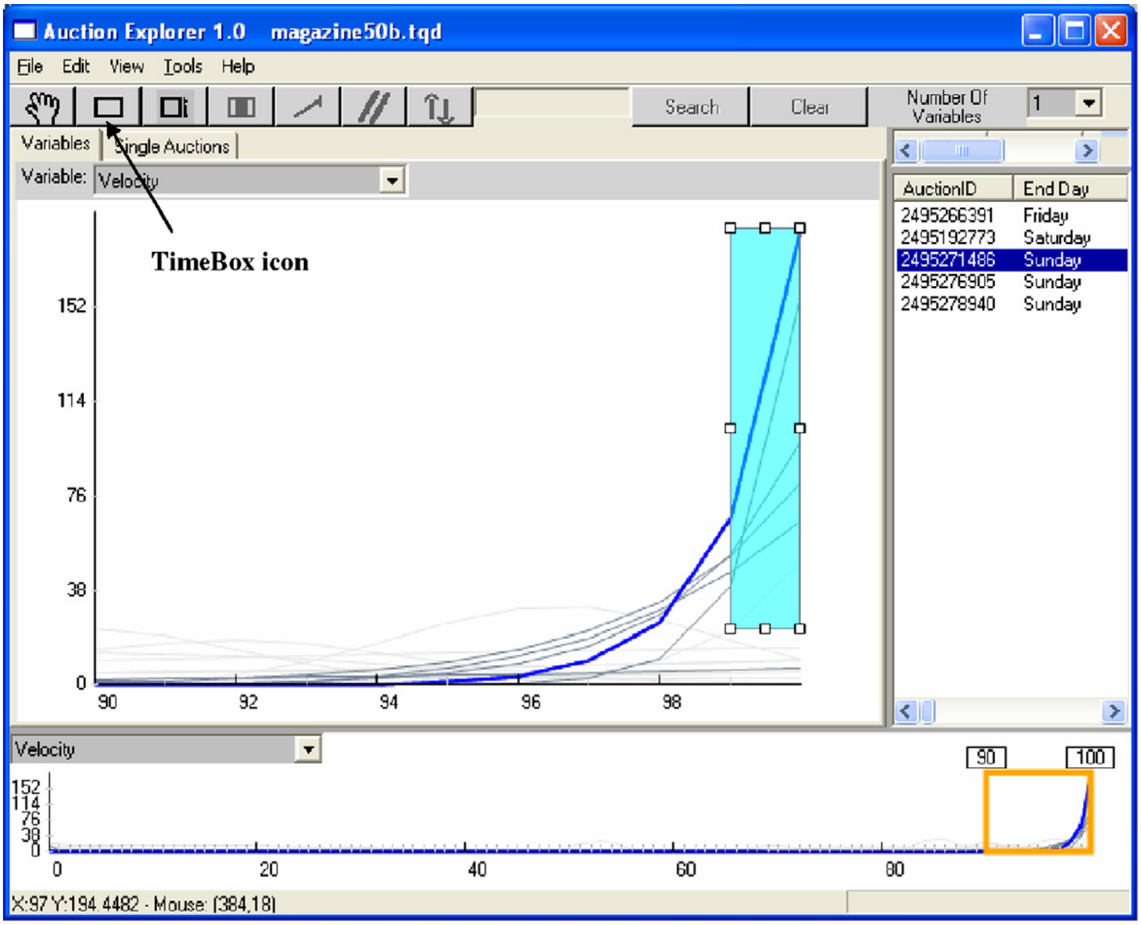


Fig. 8. Filtering auctions to show only high velocity. Only the auctions that pass through the TimeBox remain on the screen.

suggests that following a semi-structured exploration process will lead to more successful and insightful analyses. In the spirit of statistical exploratory data analysis (EDA) and data mining processes, we develop a process that can guide users in their study. The initial steps are aimed at familiarizing users with the dataset characteristics. Then, users move to studying more complicated relationships between the auction price, auction dynamics and auction attributes. These can be hypothesis-driven or exploratory.

3.2.1. Data familiarization

1. View the main screen to learn about the general size, structure, and dimension of the dataset.
2. Explore the time series panels: compare and contrast price curves and their derivative by zooming in and out, selecting and deselecting, filtering, and searching for patterns. This gives a sense of the heterogeneity across auctions, the types of prominent patterns, the times when there is action, and outliers.
3. Study the attribute panel: examine summary statistics, sort attributes and rearrange order of columns to compare multiple attributes for subsets of auctions.

For the magazines' dataset, the main screen (Fig. 5) shows the price range for the items. Users can change the log-scale of the price curves to $ to get a dollar range. They see several auctions with very high end-of-auction dynamics. Zooming in reveals that the majority of auctions actually have relatively slow dynamics. Examining the attributes users see that most auctions are 7 days long, and all auctions are in US currency, except for one in British pounds. The busiest auction has 29 bids, while the quietest auction has 4 bids. No auctions have a reserve price. One auction is an outlier in terms of ending time, which happens to be the one in British pounds (this explains why the ending time is different; it is because the seller is in a different time zone). Almost half of the auctions have no shipping cost specified. Users also see that there are no missing values for these auctions.

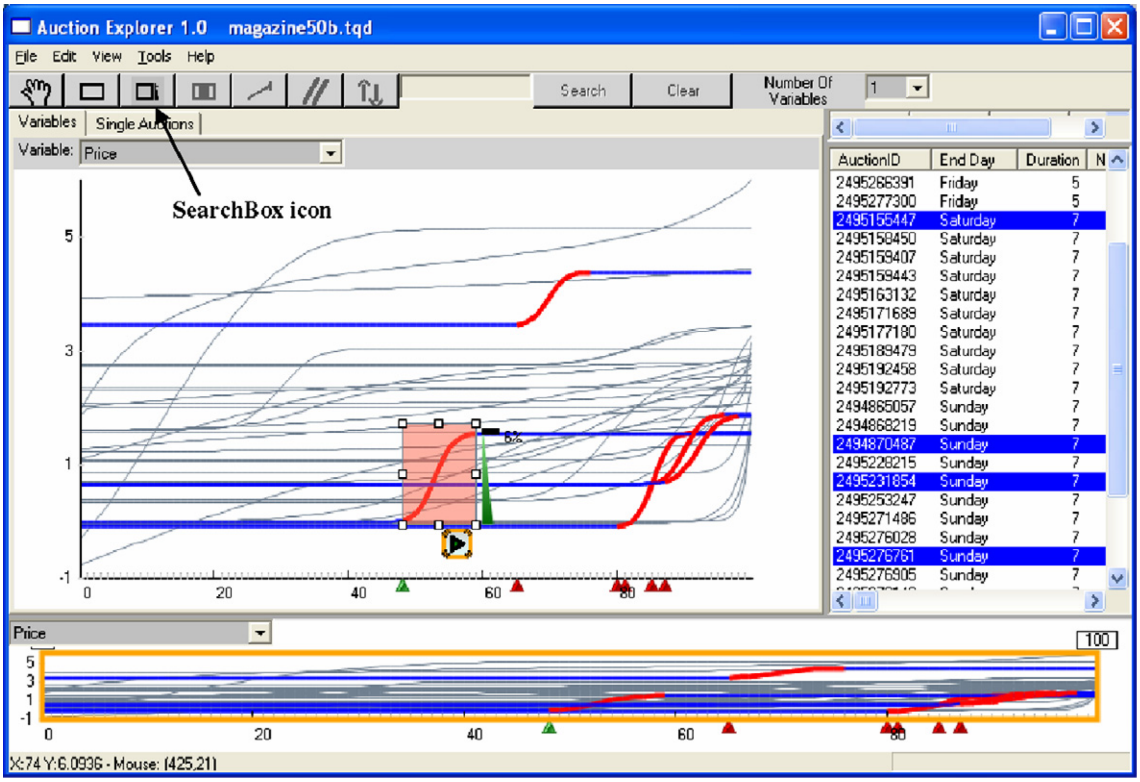


Fig. 9. Using a SearchBox to mark a pattern of interest, users can find similar patterns in the data. Here, a sharp increase in price. Such increases are found at other time periods, in different auctions.

3.2.2. *Visual search for patterns and relationships in subsets of auctions*

The next step involves exploring the relationship *between* the time series and their attributes. There are two different strategies of doing this: one is an exploratory regression-like approach, where users compare auction attributes and see how similarities (and differences) manifest themselves in the price curves and associated dynamics. The second strategy is an exploratory curve-clustering-like approach: users start by comparing price/dynamics curves, and then search for common attributes among auctions with similar curves. Clearly, these two interactive techniques are not meant to replace regression or clustering but they may prove useful when used in conjunction with analytical methods. Finding relationships by visual exploration is not confirmatory, and should be followed by formal statistical testing. The exploration is completely model-free, does not make any distributional assumptions, and cannot estimate sampling error. Thus, unlike formal models such as linear regression models, *AuctionExplorer* cannot assess whether a relationship is specific to the data sample or can be generalized to the population of interest. The exploratory regression-like and clustering-like strategies can be described as follows:

4. Select a subset of similar-attribute auctions and compare their price curves and price dynamics. Do these similar auctions have similar price dynamics? For instance, do magazines that are auctioned for 7 days, ending on a Sunday, and starting at $0.01 have similar price acceleration towards the auction end?
5. Select a subset of similar curves (with respect to price, velocity, or acceleration). This can be done by direct selection, filter boxes, or search boxes. For this subset compare their attributes by examining the summary statistics and by sorting the attribute columns. For instance, do auctions that have high acceleration near the end often end late at night?

The order of steps 4 and 5 can be reversed.

3.2.3. *Examining hypotheses about patterns and relationships*

Steps 4 and 5 can be tailored to hypotheses of interest, where the direction of the hypothesis determines which of the two strategies should be used: for

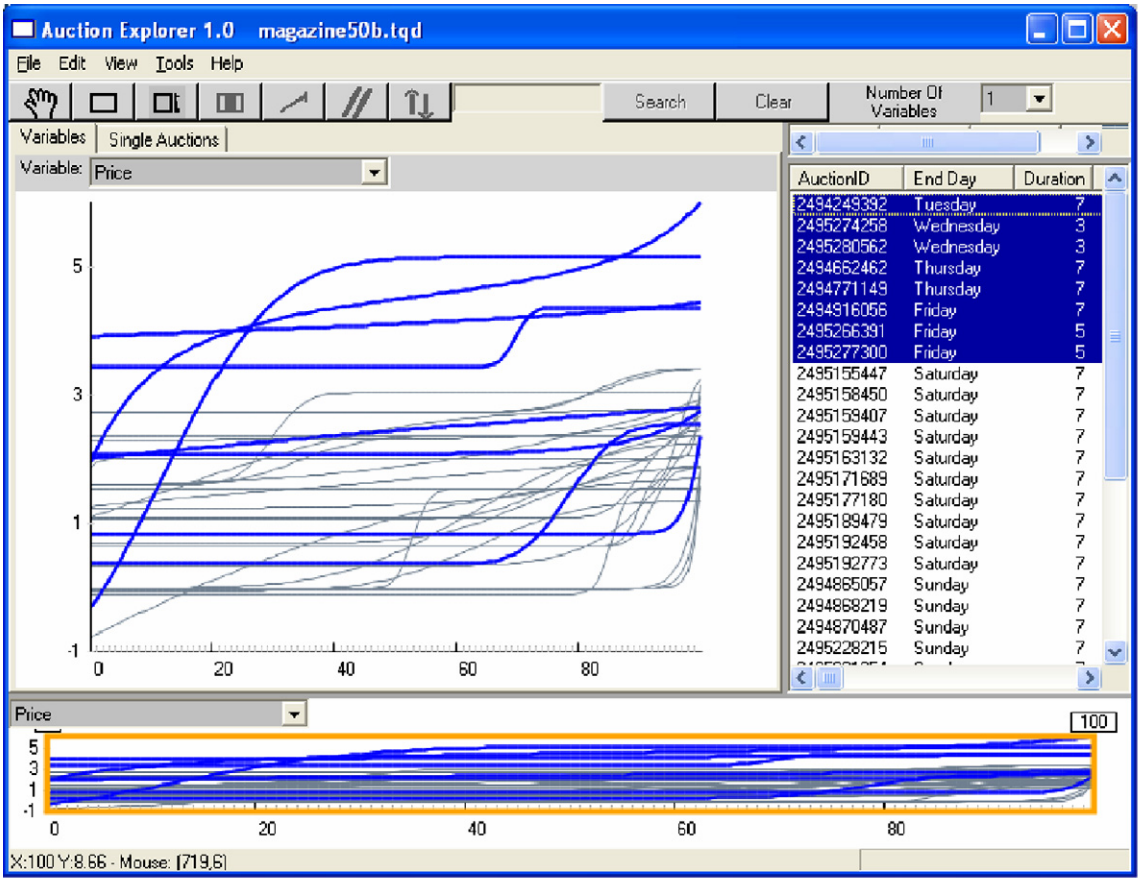


Fig. 10. Comparing prices of weekday- vs. weekend-ending auctions.

hypotheses suggesting that certain attributes lead to price dynamics, step 4 should be used. This is regression-like, where the attributes are the independent variables and the price curves are the response. An example is the combined effect of seller rating and opening price on price acceleration at the auction start. A formal statistical method for performing such an analysis is functional regression analysis where the dependent variable is a curve rather than a scalar [5,12,21,24]. Alternatively, for hypotheses regarding types of price dynamics, or regarding times of heightened price dynamics, step 5 (curve clustering strategy) is more suitable. Here the attributes are used to derive further insights about the results. One example is determining the prevalence of auctions with end-of-auction acceleration. This is related to the known phenomenon of bid sniping in eBay auctions, where intense bidding activity is observed at the last moments of the auction [22,24,25]. Another example is visual "clustering" of auctions according to their accelerating curves, and comparing the attributes of the resulting "clusters". A formal method for performing this analysis is functional curve clustering, e.g., [16]. To illustrate these exploration strategies, we offer a few hypotheses for the magazine dataset.

Hypothesis 1. *Sellers on eBay believe that ending prices are higher for auctions that close on weekends due to the increased throughput.*

Users can look for evidence for this hypothesis by examining the closing day attribute. Sorting auctions by this attribute reveals that most of the auctions in this dataset ended on weekends and only 8 auctions ended on a weekday (Fig. 10). This conforms to Hypothesis 1. However, to find out whether this belief is well grounded, users compare the price curves of weekday ending auctions (blue) to weekend ending auctions (grey). They find that the highest closing prices are actually achieved on weekday-ending auctions! This, of course, does not prove causality and is not necessarily statistically significant, but it introduces a surprising finding that should be further investigated using statistical testing.

Hypothesis 2. *Sellers that sell items in multiple auctions tend to use a consistent auction design.*

To find sellers with multiple auctions we sort by seller id (Fig. 11). In our dataset there are three sellers with multiple auctions. They all selected the same configuration for their multiple auctions (shipping price, duration, closing day, no reserve price, and opening bid). The most experienced seller even scheduled his/her three auctions to close exactly 15 min apart from each other.

One of these sellers is *woodart631*, whose two auctions are for complementary issues of the same magazine. In addition to providing further evidence for Hypothesis 2, we can also explore the effect of this seller's strategy on the auction outcome. We see that the price curves and dynamics in both auctions are identical even though the numbers of bids were different (Fig. 12). Moreover, the winner in both auctions is the same person (*bigbill2424*) who pays the same price in both auctions! Of course this might be anecdotal and statistically insignificant in such a small dataset, but in general, a finding of this type can imply to sellers and to auction houses like eBay that offering bundling can be an attractive feature, as it guarantees winning the complete set of items and can also reduce shipping costs.

In general, examining transactions by the same seller or buyer can be used for detecting possible fraudulent behavior. For instance, Ref. [18] look for indications of shilling, the act of running up the bid by the seller by using a disguised identity. One behavior that raises suspicion is a bidder who participates in many auctions by a small number of sellers and never wins. To look for suspicious auctions, *AuctionExplorer* users can search for bidders who manifest this kind of behavior, and compare the seller usernames in those auctions.

Hypothesis 3. *Last moment price dynamics are associated with high closing prices*.

The price dynamics appear to vary within our magazine auctions dataset. We use TimeBoxes (Fig. 8) to filter auctions with high velocities towards the auction end. Looking at the price curves that correspond to these auctions, we find that they actually are the auctions that closed with relatively low prices! This is in contrast to Hypothesis 3, but again, whether this is statistically

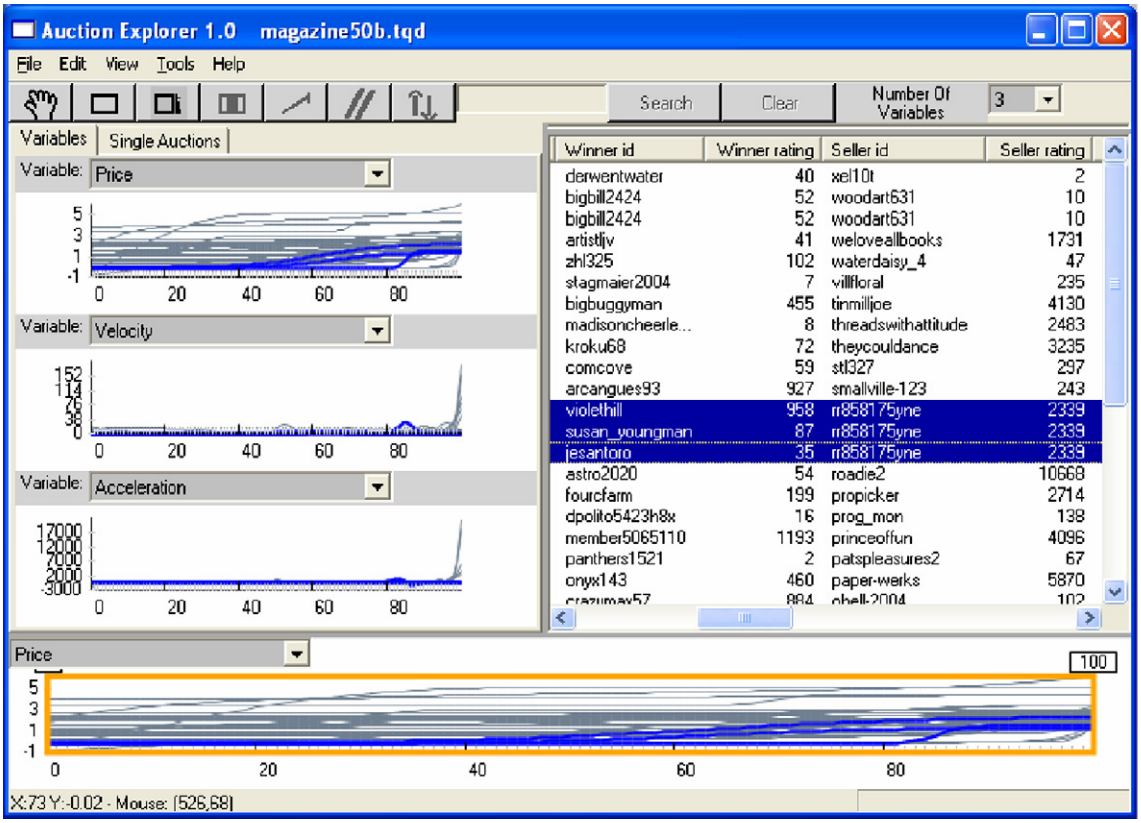


Fig. 11. Sorting by seller id to identify sellers with multiple auctions. The highlighted auctions are for the same seller (*rr858175yne*).

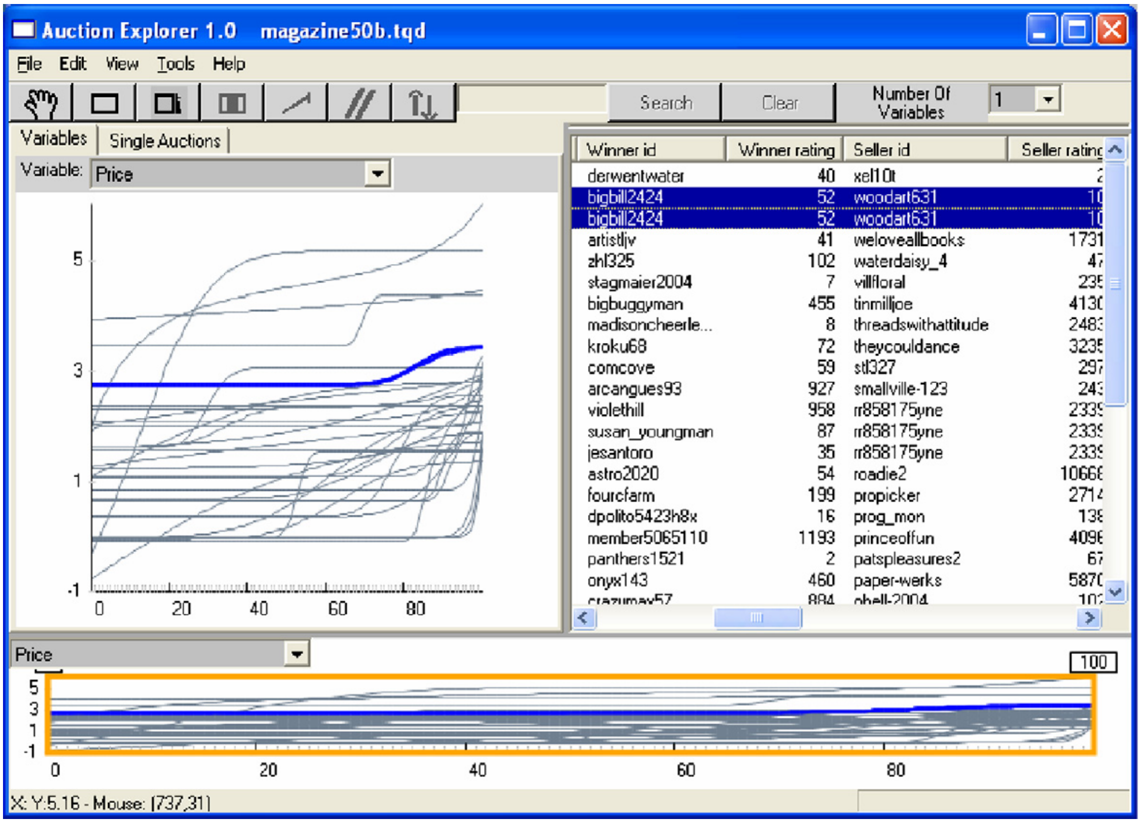


Fig. 12. Overlapping curves for two very similar auctions by seller *woodart*.

significant should be evaluated using a formal statistical model. It appears that these auctions had a too-late start, which might have prevented them from achieving high prices. Alternatively, this could result from low-priced auctions that were discovered late and won by opportunists.

4. Example datasets and performance measurements of TimeSearcher

The following are some of the datasets and their attributes that we visualized using TimeSearcher. We measured performance using the following five sample "analysis" tasks: shrinking the overview box by half, dragging a timebox, resizing the application, and selecting an item while the detailList is or is not synchronized. To measure performance we instrumented TimeSearcher with automated logging code. Each task was executed five times and to reduce the effect of the small variations in time, the average is reported (Table 1). The five "analysis tasks" described above were chosen to represent expected system usage. The first task, shrinking the overview box by half, is computationally intensive, so we wanted to verify that execution times were reasonable. TimeBox dragging is a frequent task and its time was rapid. Resizing the application is a rare task that is very computationally intensive, but its performance times were still acceptable. Selecting items, whether the detailList is synchronized or not, is a common task, so reasonable performance is important. The two tasks of application resizing and item selection while the detailList is not synchronized took approximately 3 s with 10,080 time points, a longer but acceptable delay. If larger time series are to be studied, then further code optimization efforts will be needed (Table 2).

The measurements indicate that an increase in the number of items (auctions) is as costly as an increase in the number of time points. As the number of items increase, task duration usually increases linearly. This can be observed by comparing eBay and eBay X2 datasets, as well as Magazine and Books, where both only differ in terms of the number of items. Similarly, an increase in the number of time points causes a linear increase in the duration of the tasks. This can be observed by comparing Palm 5 min, 3 min and 1 min

Table 1
Performance measurements of 5 sample tasks on 10 datasets

Dataset name	# of items	# of displayed variables	# of variables	# of time points	Task 1 (seconds)	Task 2 (seconds)	Task 3 (seconds)	Task 4 (seconds)	Task 5 (seconds)
eBay	1009	3	4	100	0.23	0.15	0.55	0.27	0.29
eBay x2	2018	3	4	100	0.48	0.26	1.12	0.50	0.52
Palm (5-min sampled)	158	1	1	2017	0.32	0.27	0.66	0.39	0.69
Palm (3-min sampled)	158	1	1	3360	0.52	0.38	1.05	0.62	1.14
Palm (1-min sampled)	158	1	1	10080	1.59	0.99	2.96	1.84	3.33
Palm (smooth curves)	158	3	3	71	0.05	0.05	0.16	0.05	0.08
Magazine	34	3	3	100	0.03	0.02	0.11	0.03	0.05
Books	227	3	3	101	0.06	0.06	0.18	0.06	0.10
Books	227	2	3	101	0.06	0.03	0.16	0.06	0.09
Books	227	1	3	101	0.06	0.05	0.13	0.04	0.07

datasets. While consecutive missing values might increase performance (due to decreased number of processing requirements for certain subtasks such as drawing), sporadic missing values may introduce overhead. In addition, the linear relationship may not be apparent when comparing small datasets such as Magazine and Books. This is because the constant factor might dominate the linear component in performance measurements. The tasks of shrinking the overview box and selecting an item while the detailList is not synchronized are believed to be the most heavily used tasks. Resizing the application occurs occasionally, while dragging a timebox occurs when users explore the dataset by visual TimeBox filtering. The measurements for the sample tasks are provided in seconds and varied with repeated runs by approximately 5% due to operating system factors, such as memory caching. The tasks were simulated by a program to prevent measuring user variation, such as difference between a slow and a fast dragging movement. The tasks were performed on Windows XP platform, where .NET Framework 1.1 was installed on a Dell 8400 with 2.4 GHz Intel Pentium processor and 1 GB installed RAM.

Table 2
Description of datasets

Dataset Name	Description
eBay	A variety of eBay auctions on items across multiple categories, in 3 currencies
Palm	eBay auctions on 7-day auctions for Palm M515 PDA. The versions differ according to how the curves were created from the bid history (by sampling at different intervals, and smoothing)
Books	eBay auctions on items sold in the books category
Magazines	A subset of the books data, within the subcategory of magazines

5. Conclusion and future work

AuctionExplorer is a suite of tools that support the collection, representation, and interactive exploration of datasets with auction attributes (cross-sectional data) and bid histories (time series). *AuctionExplorer* enables users to cope with the abundance of information that is contained in these two information structures and in their relationships. With a range of operations for manipulating attributes and time series, the visualization tool can be used by both academic researchers and practitioners (i.e. eBay bidders and sellers, and eBay itself). For instance, an academic researcher could use this tool to explore relationships between auction design, seller characteristics and price dynamics and might find that lower opening bids coupled with higher seller reputation results in a faster price acceleration at the auction end. Such an observation could then be the starting point for developing formal econometric models that confirm and explain such relationships within the grander theory of electronic markets. On the other hand, an eBay bidder who needs to choose the auction to participate in, and the time and amount to bid, could use such a tool to explore the relationship between auction design and bidder competition. A possible finding (which should then be examined more formally through statistical testing) would be that shorter auctions that close in the early morning hours attract a smaller number of competing bidders and result in lower price velocity at the end (implying a smaller danger of being “out-sniped” in the last seconds). Also, a seller, who wants to learn about design choices that induce high dynamics and high prices, could use this tool to investigate which auction design features resulted in the highest price dynamics. Higher price dynamics are associated with a higher price and thus a larger profit margin for the seller. Finally, this tool is useful for an auction house, which can benefit from monitoring their transactions for purposes of improving pricing and design of the auction site and even for detecting fraud.

Future possible enhancements of *AuctionExplorer* that would further assist users in the decision making process include developing a "front end" that would allow users to answer specific common questions, and perhaps a statistical tool for testing whether visually detected patterns are significant or random noise.

Although we designed *AuctionExplorer* for online auction data, it can be used for a much more general class of applications. The online environment leads to the coupling of time series with cross-sectional data in many applications. One example is user ratings of products or services, such as movie ratings on Yahoo! [11] and product rating on Amazon [9] where each record contains a history of ratings as well as attribute data. Another example is publicly available data on open-source software development projects, e.g., on freshmeat.com, [26], where the time series is comprised of the size and complexity of software releases over time, and attributes are project type, number of subscribers, etc. Other examples include the price of a set of products, such as books, at multiple e-tailers over time, prices of airline tickets on different carriers on a certain route, etc. Analysts with tools that can assist in understanding and exploring the combination of time series with attribute data will have a distinct advantage.

Acknowledgements

The authors thank Ravi Bapna and four anonymous referees for their helpful comments. Partial data for this project was supplied by CIDRIS — the Center for Internet Data and Research Intelligence Services, University of Connecticut. This research was partially funded by the NSF grant DMI-0205489.

References

[1] G.N. Allen, S.T. March, Writing electronic commerce data collection agents: a tutorial using visual basic 6.0, Proceedings of the 2000 International Conference on Information Systems (ICIS), Brisbane, Australia, 2000.

[2] A. Aris, B. Shneiderman, C. Plaisant, G. Shmueli, W. Jank, Representing unevenly-spaced time series data for visualization and interactive exploration, Proceedings of Interact 2005, Springer, Berlin, 2005, pp. 835–846.

[3] P. Bajari, A. Hortacsu, The winner's curse, reserve prices and endogenous entry: empirical insights from eBay auctions, Rand Journal of Economics 3 (2) (2003) 329–355.

[4] R. Bapna, P. Goes, A. Gupta, Y. Jin, User heterogeneity and its impact on electronic auction market design: an empirical exploration, MIS Quarterly 28 (1) (Mar. 2004) 21–43.

[5] R. Bapna, W. Jank, G. Shmueli, Price Formation and its Dynamics in Online Auctions, working paper, Smith School of Business, University of Maryland (2005).

[6] R. Bapna, W. Jank, G. Shmueli, Consumer Surplus in Online Auctions, Working paper, University of Connecticut (2005).

[7] S. Borle, P. Boatwright, and J.B. Kadane, The Timing of Bid Placement and Extent of Multiple Bidding: An Empirical Investigation Using eBay Online Auctions, Working paper, Rice University (2005).

[8] P. Buono, A. Aris, C. Plaisant, A. Khella, B. Shneiderman, Interactive pattern search in time series, Proceedings of Conference on Visualization and Data Analysis, SPIE, Washington, DC, 2005, pp. 175–186.

[9] J.A. Chevalier and D. Mayzlin, The effect of word of mouth on sales: online book reviews, Yale SOM Working Paper No's. ES-28 and MK-15 (Aug 6, 2003).

[10] C.N. Dellarocas, Analyzing the Economic Efficiency of eBay-like Online Reputation Reporting Mechanisms, Proceedings of the 3rd ACM Conference on Electronic Commerce, Tampa, FL, October 14–16, 2001.

[11] C.N. Dellarocas, N. Awad, X. Zhang, Using online reviews as a proxy of word-of-mouth for motion picture revenue forecasting, Working paper, MIT (May 10, 2004).

[12] J.J. Faraway, Regression analysis for a functional response, Technometrics 39 (1997) 254–261.

[13] H. Hochheiser. Interactive Graphical Querying of Time Series and Linear Sequence Data Sets, Ph.D. Dissertation, University of Maryland, Department of Computer Science (May 2003).

[14] H. Hochheiser, B. Shneiderman, Dynamic query tools for time series data sets, timebox widgets for interactive exploration, Information Visualization 3 (1) (Mar. 2004) 1–18.

[15] J.A. Jacko, A. Sears, M. Beaudouin-Lafon, R.J.K. Jacob (Eds.), The Dynamics of Mass Online Marketplaces: A Case Study on an Online Auction, Proceedings of the CHI 2001 Human Factors in Computing Systems, Seattle, WA, March 31–April 5, 2001, vol. 3 (1), 2001, pp. 317–324.

[16] W. Jank, G. Shmueli, Profiling Price Dynamics in Online Auctions Using Curve Clustering, Working Paper, Smith School of Business, University of Maryland (2005).

[17] W. Jank, G. Shmueli, Functional data analysis in electronic commerce research, Working paper, Smith School of Business, University of Maryland (2005).

[18] R.J. Kauffman, C.A. Wood, Running up the bid, Modeling Seller Opportunism in Internet Auctions, Proceedings of the AMCIS 2000 Conference, Long Beach CA, 2000.

[19] R.J. Kauffman, S.T. March, and C.A. Wood. Agent Sophistication: Design Aspects for Data-collecting Agents, to appear in International Journal of Intelligent Systems in Accounting, Finance, and Management.

[20] D. Lucking-Reiley, Auctions on the Internet: what's being auctioned and how? Journal of Industrial Economics 48 (3) (2000) 227–252.

[21] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer-Verlag, New York, 2005.

[22] A.E. Roth, A. Ockenfels, Last-minute bidding and the rules for ending second-price auctions: evidence from eBay and amazon auctions on the Internet, The American Economic Review 92 (4) (2002) 1093–1103.

[23] G. Shmueli, W. Jank, Visualizing online auctions, Journal of Computational and Graphical Statistics 14 (2) (2005) 299–319.

[24] G. Shmueli, W. Jank, Modeling the dynamics of online auctions: a modern statistical approach, in: R. Kauffman, P. Tallon (Eds.),

Economics, Information Systems and Ecommerce Research: II. Advanced Empirical Methods, part of Advances in Management Information Systems Series, M.E. Sharpe, Armonk, NY, in press.

[25] G. Shmueli, R. Russo, and W. Jank. Modeling Bid Arrivals in Online Auctions, Working Paper, University of Maryland (2004), www.smith.umd.edu/ceme/statistics/BidArrivals_ Jan04.pdf.

[26] K.J. Stewart, T. Ammeter, L. Maruping, A preliminary analysis of the influences of licensing and organizational sponsorship on success in open source projects, Proceedings of the 38th Hawaii International Conference on System Sciences, Jan. 2005.

[27] Y. Vakrat, A. Seidmann, Implications of the bidders' arrival process on the design of online auctions, Proceedings of the Hawaii International Conference on System Sciences, 2000.

Galit Shmueli is Assistant Professor of Management Science and Statistics at the Smith School of Business, University of Maryland. She holds a PhD in statistics from the Technion - Israel Institute of Technology (2000). Her research interests lie in developing and applying statistical methods and models for unique data structures. These include visualization, non-parametric methods, advanced monitoring algorithms, and specialized flexible models for discrete data. Her two main areas of application are in electronic commerce and in biosurveillance, where new data structures give rise to statistical challenges.

Wolfgang Jank is Assistant Professor of Management Science and Statistics in the Smith School of Business at the University of Maryland. He obtained his PhD in Statistics from the University of Florida in 2001. His methodological research interests center around computational statistics, functional data analysis, methods for spatial and temporal data, Monte Carlo methodology, stochastic optimization and information visualization. He is interested in applications in electronic commerce, marketing, operations management, and aviation.

Ben Shneiderman (http://www.cs.umd.edu/~ben) is a Professor in the Department of Computer Science, the Founding Director (1983–2000) of the Human-Computer Interaction Laboratory (http://www.cs.umd.edu/hcil/), and a Member of the Institutes for Advanced Computer Studies & for Systems Research, all at the University of Maryland at College Park. He was elected as a Fellow of the Association for Computing (ACM) in 1997 and a Fellow of the American Association for the Advancement of Science (AAAS) in 2001. He received the ACM SIGCHI Lifetime Achievement Award in 2001. Ben is the author of Software Psychology: Human Factors in Computer and Information Systems (1980) and Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th ed. 2004). He pioneered the highlighted textual link in 1983, and it became part of Hyperties, a precursor to the web. His move into information visualization helped spawn the successful company Spotfire. He is a technical advisor for the HiveGroup and ILOG. With S. Card and J. Mackinlay, he co-authored Readings in Information Visualization: Using Vision to Think (1999). His recent books include Leonardo's Laptop: Human Needs and the New Computing Technologies (MIT Press) and with B. Bederson, The Craft of Information Visualization (Morgan Kaufmann).

Dr. Catherine Plaisant is Associate Research Scientist at the Human-Computer Interaction Laboratory of the University of Maryland Institute for Advanced Computer Studies. She earned a Doctorat d'Ingenieur degree in France in 1982. In 1987, she joined Professor Ben Shneiderman at the Human-Computer Interaction Laboratory. She enjoys most working with multidisciplinary teams on designing and evaluating new interface technologies that are useable and useful. Her research contributions range from focused user interaction techniques (e.g. Excentric Labeling) to innovative visualizations (such as LifeLines for personal records or SpaceTree for hierarchical data exploration) and interactive search interface techniques such as Query Previews. Those interaction techniques have been carefully validated with user studies and are finding applications in industry and government information systems and digital libraries. She has written over 80 refereed technical publications on the subjects of information visualization, digital libraries, universal access, image browsing, input devices, online help, home automation, network management, telemedicine, etc. She recently co-authored with Ben Shneiderman the 4th Edition of Designing the User Interface, one of the major books on the topic of Human-Computer Interaction.

Aleks Aris is a PhD student at the Computer Science Department and a member of the Human-Computer Interaction Laboratory (HCIL) at the University of Maryland, College Park. He was the major developer of recent versions of Treemap and from early to latest versions of TimeSearcher 2, which he recently extended to AuctionExplorer. He also worked on the MyLifeBits project of Microsoft Research during the summer of 2004. His current research interests are directed towards network visualization. These include increasing the effectiveness of the visual display and designing the user interface to support tasks for users who interact with and explore networks.