FacetLens: Exposing Trends and Relationships to Support Sensemaking within Faceted Datasets

Bongshin Lee, Greg Smith, George G. Robertson, Mary Czerwinski, Desney S. Tan
Microsoft Research
One Microsoft Way, Redmond, WA 98052
{bongshin, gregsmi, ggr, marycz, desney}@microsoft.com

ABSTRACT
Previous research has shown that faceted browsing is effective and enjoyable in searching and browsing large collections of data. In this work, we explore the efficacy of interactive visualization systems in supporting exploration and sensemaking within faceted datasets. To do this, we developed an interactive visualization system called FacetLens, which exposes trends and relationships within faceted datasets. FacetLens implements linear facets to enable users not only to identify trends but also to easily compare several trends simultaneously. Furthermore, it offers pivot operations to allow users to navigate the faceted dataset using relationships between items. We evaluate the utility of the system through a description of insights gained while experts used the system to explore the CHI publication repository as well as a database of funding grant data, and report a formative user study that identified usability issues.

Author Keywords
Facets, interactive visualization, trends, relationships, sensemaking.

ACM Classification Keywords
H.5.2 [User Interfaces]: Graphical User Interface.

INTRODUCTION
We are generating data at an increasingly prolific rate. Regardless of the domain, be it conference publication data, research spending, or even personal digital artifacts, we would often like to explore, make sense of, and derive insights from large numbers of collected digital items.

Traditionally, computer users have used one of several strategies to do this. The simplest and most popular strategy is direct search. In direct search, users issue explicit queries to find specific items within the dataset. Unfortunately, this does not explicitly support exploring relationships that exist between data items. Furthermore, it fails when users do not have exact keywords in mind. Navigational search, another commonly used strategy, imposes a hierarchical classification on data so that users can browse the information by iteratively narrowing the scope in a predetermined order. Unfortunately, the fixed ordering of categories does not expose the richness contained within multi-dimensional datasets. Moreover, the selection and maintenance of the classification strategy is itself an extremely difficult job. Recently, researchers have explored faceted search, which supports the assignment of multiple unordered classifications to each item within the system. Since faceted classifications are complementary and independent, users have much more freedom to navigate, narrow down the data, and pivot across various dimensions to explore relationships.

While previous work has focused on search and browse tasks, we believe there are large opportunities for exploring the efficacy of interactive visualization systems in supporting exploration and sensemaking within faceted datasets. To do this, users need to detect and understand meaningful trends and relationships that allow them to describe the past, analyze the present, and predict the future [6]. While current faceted systems go a long way in providing such utility, we present several ways in which we can improve upon current methodologies.

In this paper, we present FacetLens, an interactive visualization system designed to expose trends and relationships within faceted datasets. FacetLens extends previous faceted systems in two ways. First, in addition to more traditional facet types such as single-value, multi-value, and hierarchical, FacetLens implements *linear facets*. While attribute values are intrinsically categorical, linear facets permit the visual representation of order within a facet in a way that allows data trends such as temporal relationships to be preserved and exposed. Second, it provides users with the ability to *pivot* between related facets at any point during the exploration. This is important because it allows users to maintain a sense of context while they quickly and efficiently explore various areas within the dataset.

The specific contributions of this paper are:

1. Description of the FacetLens interface, focusing on how linear facets, which allow users to compose and compare bar histograms and pivoting, expose trends and relationships and support sensemaking tasks.

2. Proof of utility through a description of insights gained while experts used the system to explore two datasets: the CHI publication repository running from 1982 through 2004; and a dataset from the Organization for Economic Cooperation and Development (OECD) containing records representing nearly a million aid grants made between countries over the last 32 years.
3. Reports from a formative usability study showing general success with novice users, but more importantly identifying areas for improvement.

We conclude the paper with discussion of the results and limitations, implications for design, and proposals for possible future work.

RELATED WORK

As suggested by its name, the creation of FacetLens was influenced by two pieces of previous work, PaperLens [7] and FacetMap [12]. PaperLens is a carefully crafted interface that facilitates targeted question-answering within publication datasets. For example, the system is tailored to reveal the evolution of key topics, relationships between authors, as well as the most frequently referenced papers and authors. However, because of its design specificity, PaperLens does not support discovery and exploration of arbitrary questions well. FacetMap, on the other hand, is a faceted browser designed for data and display scalability as well as to be question agnostic. However, since FacetMap was designed primarily for simple search and browse tasks, it does not support exploring trends or relationships between data items, hence limiting the set of insights that can be derived. In this section, we review research on faceted classification systems, trend visualization, and other techniques that motivate our approach.

Faceted Classification Systems

In faceted classification systems, the attributes of the items are grouped into multiple orthogonal categories called “facets.” For example, a database of books might have a Year facet to group together publications from the same year, and a Genre facet exposing attribute values like “fiction,” “history,” or “education.” Providing multiple ways to reach items by presenting several facets simultaneously alleviates the drawbacks of any single categorization scheme.

The Flamenco system [18] was an early demonstration that faceted browsing, combined with comprehensive dynamic query capabilities, can provide a more effective and enjoyable searching and browsing user experience than keyword search or pure categorization. Relation Browser++ [19] presents facets and their top-level values simultaneously, with embedded bar graphics to visualize the value distributions and mouse-over effects for query previews. It is adaptable to different datasets, but limited in the amount of metadata allowed before navigation becomes burdensome. FacetMap [12] provides a dynamic, graphical faceted browsing experience that can be applied to a wide range of datasets without re-authoring the interface, but suffers limitations in its ability to represent inter-item relationships and non-hierarchical facets.

The Colon classification [10] is an example of faceted classification used in the real world. Faceted classification has become very popular in recent years, especially on the web, and many commercial online shopping experiences, such as Audible (audible.com) and Shopping (shopping.com) now provide faceted browsing.

Exploratory Data Analysis

Information visualization supports exploratory data analysis, which helps people gain meaningful insights by revealing underlying patterns and relationships. High-end data exploration tools such as Spotfire [1] and Polaris [14] provide high levels of customization and control to users. InfoZoom [13] also gives full control to users to interactively explore the dataset without providing any automatic data mining results. These tools are often targeted for expert users who not only are proficient with the tool but also have good knowledge in the domain. On the other hand, Tableau [8] tries to provide good default visualizations so that users can concentrate on their analysis. These systems all become difficult to use if the user is not sure where to look for patterns or how to pose the proper query. FacetLens addresses this issue with faceted browsing, which supports many paths to the same information.

Trend Visualization

Revealing a trend is a simple but powerful data summarization technique, and several visualization strategies exist to support this. The simplest and most common way is to plot a variable’s change over time on a line chart or bar chart. Many commercial tools including Excel [9] and Tableau [8] support this approach. Several systems such as PaperLens [7], NetLens [5], and IN-SPIRE [17] use multiple bar charts to show trends for a certain variable. When multiple variables are involved, the small multiples approach to a bar chart or line chart can become prohibitively space-intensive. To save screen space, other systems have tried a stacking approach. NameVoyager [15] examines a dataset of baby names over time in the United States. ThemeRiver [4] is a timeline indicating the flow of document themes. It uses width of the river to show the number of documents and the river is sub-divided by topics, which ebb and flow over time. Some previous facet-based visualizations such as Bungee View [2], Relation Browser++ [19], and InfoZoom [13] use bar graphs to show value distributions. However, those bar graphs are not coupled with any attribute values having explicit order (e.g., time). In other words, they are not designed to show changes of values over time but simply to show how many items are in each category.

Recognizing the importance of displaying trends in order to convey understanding of data, especially multi-dimensional data, Gapminder Trendalyzer [3] uses an animated bubble chart to show trends over time in three dimensions: x position, y position, and bubble size. Robertson et al. [11] proposed two alternative approaches that use static depictions of trends. They also evaluated the effectiveness of three

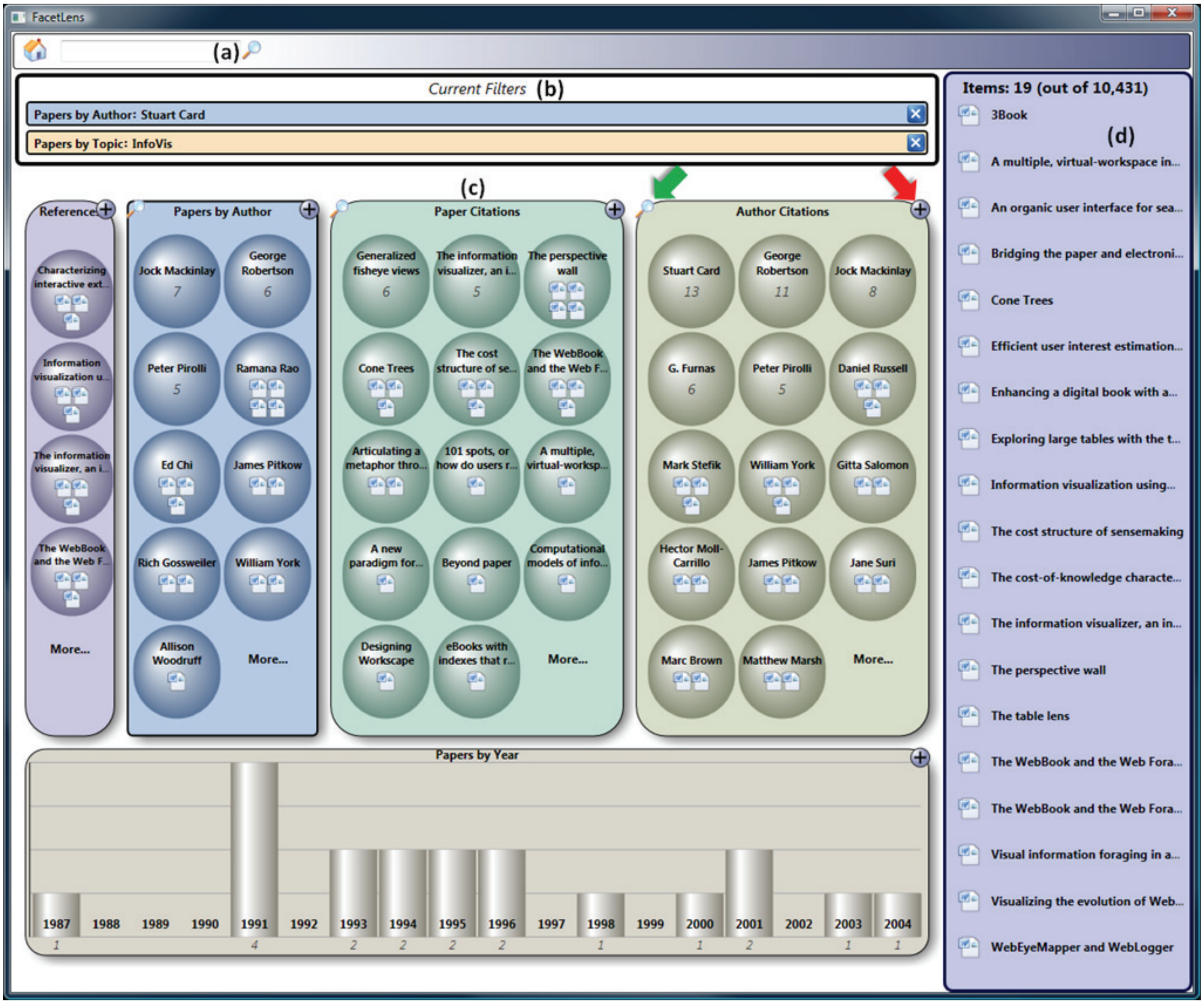

Figure 1. FacetLens consists of four areas: a) The search type-in box is used for a standard full-text item search. b) The Current Filters area at the top shows the applied filter clauses. c) The main facet area shows the currently available facets and attribute values. d) The Items area on the right shows the result set items in standard list form. The green arrow points to the button for searching within a facet, and the red arrow points to the facet flyout button. (These two arrows are added for illustration.)

visualizations for two types of activities; analysis and presentation. Using linear facets, which will be described in the next section, FacetLens allows users to explore trends at various levels of granularity: trends for the whole dataset, trends for one facet, and trends for the set of items with a specific attribute value.

FACETLENS INTERFACE

There are four main components within the FacetLens interface: the Facet area, which consumes the bulk of the interface, the Current Filters above it, the Search Bar on top, and the Item Details pane on the right (see Figure 1). We ground our discussion of the interface within a conference publication dataset, in which papers and authors form the core items within the system, and facets include authorship, topics, references, citations, and so on.

The interface starts by presenting an overview of all the information in the dataset. Attribute values are graphically grouped by facet (e.g., Authored By, References, Topics) in the Facet area, with each facet encapsulated by a uniquely colored rounded rectangular box. Each attribute value (e.g., Stuart Card, George Robertson, Jock Mackinlay) is contained within a similarly colored elliptical bubble, and each bubble also contains a count of the number of items that have that attribute.

Because an item is presented in each of its attribute value groupings across the facet space, at certain levels of detail the same item may appear simultaneously at several different locations in the interface. When the number of items in a bubble is smaller than a threshold (e.g., five), FacetLens presents miniature embedded icons for these items (e.g., see Figure 2), increasing the number of items in a given view even further. To help the user understand the correspondence across the interface between all presentation instances of a given item, FacetLens uses a brushing effect: when the cursor passes over an item, all instances of that item in view are highlighted (Figure 2).

Basic interaction with FacetLens is straightforward. Users click on attribute values to create filters that narrow the view on the data. For example, clicking on Stuart Card in the Papers By Author facet restricts the view to represent only papers that Stuart Card has authored. Filters can be combined (using a logical AND operation), so clicking on InfoVis in the Topic facet after the previous action further limits the view (Figure 1). The entire set of currently-applied filter clauses can be seen at a glance in the Current Filters area at the top (Figure 1b).

Whenever the current filter changes, we remove facets and attribute values that no longer describe any of the remaining items, and reallocate the available screen space amongst the facets and attribute values of the remaining items. For example, once the Usability value has been selected as a filter from the Papers by Topic facet and this clause is added to the Current Filter area, the Papers by Topic facet can disappear because it has no further metadata available to display, and the remaining space can be filled with further detail from facets that still contain useful metadata structure.

At each point in the interaction, items remaining within the current view are presented in the Item Details pane (Figure 1d). This is a column along the entire right-hand side of the interface containing items in standard list form, which helps users anchor the results with respect to the current scope of the dataset while they navigate and make sense of the dynamic Facet area. When the current number of items is too large to be completely displayed within this column, the top-N items are shown (according to a configurable sort or rank criterion). When the scope of the dataset is smaller, the individual item spaces grow vertically to display more embedded item detail (Figure 5). After the previous two interactions, 19 items remain, indicating that Stuart Card has authored 19 InfoVis papers within this dataset.

Clicking on the miniature icons invokes the Item Details window showing the detailed attributes of a single item. For example, Figure 3 shows the attributes of the paper Cone Trees and the author Jock Mackinlay. In this window, FacetLens also provides a button for the pivot operation, which will be described in detail later.

We believe that this paradigm of dynamic, interactive information presentation allows effective navigation through the dataset. This is powerful when answering targeted questions that are supported by the structure of the facets. However, we found that our original implementations of this paradigm fell short when trying to visualize more complex trends and relationships within or between facets. In the following subsections, we describe the two main innovations within FacetLens: linear facets that support identifying and comparing trends within facets; and pivoting, which

Figure 2. Flyout window for the "Papers by Author" facet with "Brad Myers" filter applied shows his co-authors. One particular paper is highlighted under four different co-authors on mouseover.

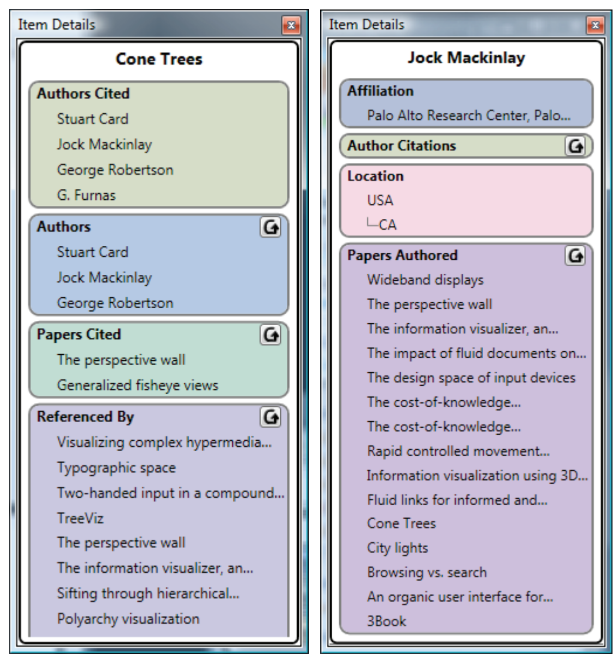

Figure 3. Two examples of the Item Details window, which presents all of an item's attribute values in detail. The pivot button (shown as a circular arrow in the upper right of certain attribute regions) is available when the attributes of the item are items themselves in the metadata space of another facet.

permits quick exploration to exploit significant relationships across the dataset. Additionally, we present dynamic screen allocation among attribute values across all facet types and an augmentation to the traditional directed search mechanism that allows users to search on attribute values.

Linear Facets: Identifying and Comparing Trends

One set of interesting insights that can be made about a large collection of data such as a conference publication dataset involves meaningful trends. For example, which topics have come and gone in a conference? What is the publication trend of an author or an institute? What is the citation pattern of an author or a paper? Unfortunately, since the focus of most faceted systems is on presenting categorical selections, many systems also end up removing important visual ordering information. Facets are treated as lists of labels on subsets of the data, and so are attribute values within the facets.

FacetLens introduces the *linear* facet in which attribute values are visually presented in an explicit order, such as time. This permits exposing rich relationships between attribute values within a facet. Within the interface, linear facets are displayed horizontally across the bottom of the Facet area. The attribute value bubbles are displayed as columns rising from a common base with a height proportional to the item count, creating the appearance of a histogram (Figure 1 and 4). They are sorted by their attribute values. In addition, missing attribute values along the linear scale of the facet are allocated empty column space to preserve the linear continuity of the horizontal axis. Since the differences between linear facets and regular facets lie only in the visual presentation, the attribute values in the linear facet are still clickable for filter application, and as space permits greater detail the columns are recursively filled in with finer-granularity histograms or items as appropriate. Linear layout of certain facets allows peripheral spotting of trends even when those facets are not part of the interaction, because, like all facets in FacetLens, linear facets react to the application of filter clauses by reallocating space to display only the filtered items. For example, if Brad Myers in the Papers By Author facet is selected as a filter, Brad Myers' publication trends across the years are readily discernible in the Papers by Year linear facet.

To further allow users to explore trends *without* applying filter clauses, FacetLens introduces a flyout window. Dragging a bubble onto a facet in the main window (rather than clicking on it) automatically invokes the flyout to show the dragged bubble's items grouped according to the attribute values of the target facet. For example, dragging Stuart Card in the Papers By Author facet onto the linear facet Papers by Year will show Stuart's papers broken down by year of publication in a flyout window, which will reveal Stuart's publication trends. However, because the filter clause is unaffected by the drag/drop operation (this is a sandboxed operation), users can compare trends among multiple disjoint item subsets by dragging additional attribute value bubbles onto the same flyout. For example, additionally dragging Ben Shneiderman onto the flyout window will show Ben's publication trends underneath Stuart's (Figure 4).

Pivoting: Navigating between Related Items

Since filtering is the core and sometimes only interaction technique in traditional faceted browsers, users often find themselves unable to proceed with potentially useful exploration once they have exhausted all filters. When this occurs, they are left with a set of items, from which they must either remove filters and try new ones, or start a brand new search. This is true even though there are often complex and interesting relationships between items.

To enable users to navigate further into related items whether or not filters have been exhausted, FacetLens offers a new operation we call a *pivot*. Pivots are a means to reset the view to show related items, enabled by allowing items to be attributes of other items. For example, in the Papers by Author facet, author items are attributes of paper

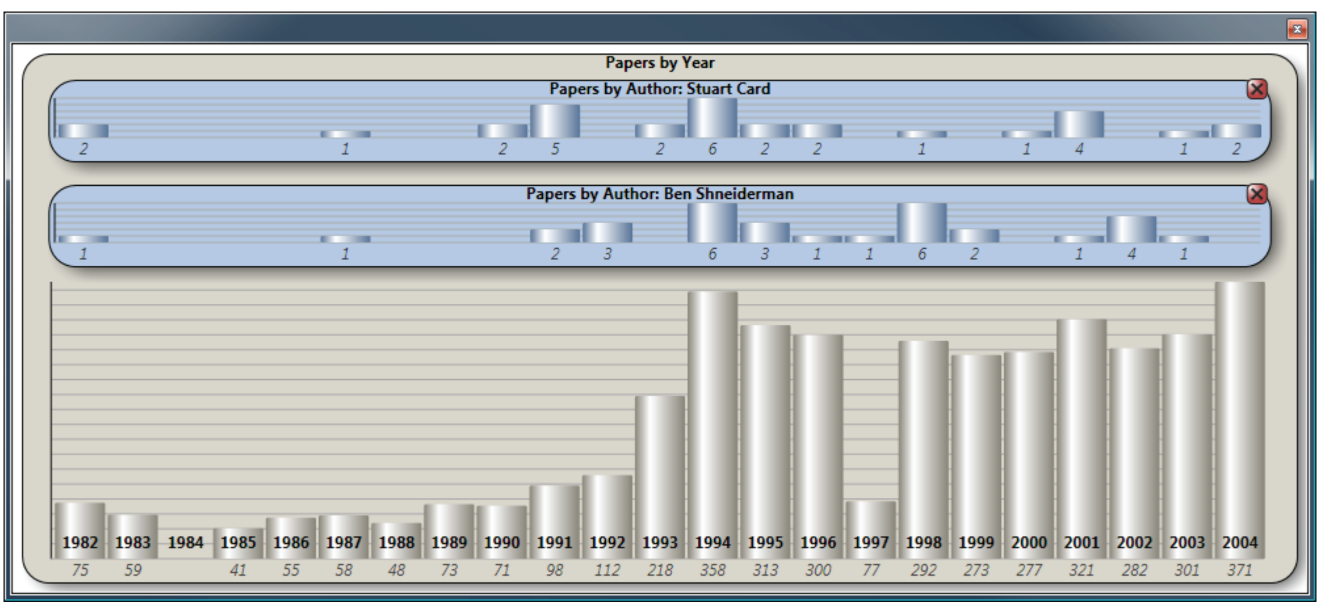

Figure 4. The "Papers by Year" linear facet with two other facets that were dragged onto "Papers by Year" for comparison.

items, but each author attribute bubble also represents an author item in the dataset (Figure 1 and 2).

The pivot operation is offered in the Items Details window, where a related item is listed as an attribute. For example, in the Item Details window for the Cone Trees paper, users can pivot to the authors view to learn more about the three authors of the paper by clicking on the circular arrow button in the Authors attribute subpanel (Figure 3). This resets the global filter to Authors by Paper: Cone Trees, where the three authors (Robertson, Mackinlay, and Card) are now the only items in scope and the Facet area displays their metadata accordingly (Figure 5). If users want to know more about the papers written by one of the authors of the Cone Trees paper, they can pivot back into a papers view from this authors view. For example, they can pivot to show papers written by Jock Mackinlay from the Item Details pane (Figure 5) or from the Item Details window for the author, Jock Mackinlay (Figure 3).

The pivot operation can also be performed within an item type. For example, users can pivot to the papers view to learn about the papers that referenced the Cone Trees paper by clicking on the circular arrow button in the Referenced By attribute subpanel (Figure 3).

The item scoping path for this navigation is Papers (papers written by Stuart Card) → Authors (authors of the Cone Trees paper, one of the Stuart Card's papers) → Papers (papers written by Jock Mackinlay, one of the authors of the Cone Trees paper) → Papers (papers referenced the Cone Trees paper, one of the Jock Mackinlay's papers). Note that users can navigate further by deciding what to follow while they are exploring the dataset.

Dynamic Space Allocation: Suggesting the Next Step and Revealing Relationships

In most faceted visualizations, the interface is carefully authored to expose the most appropriate metadata for narrowing the dataset in the initial view. As metadata filters are applied, the distribution of metadata among the remaining items inevitably changes, and the original authored allocation of space to specific metadata may no longer be relevant or appropriate.

As the basic interaction descriptions above make clear, FacetLens seeks to avoid this issue by constantly re-evaluating the distribution of metadata among the current set of items, and mapping more visual space to regions of metadata that contain more items in the current scope. The count of items is used as an importance metric among facets and attribute values. When applied consistently across the entire visualization, this strategy has three beneficial effects: First, users quickly develop a sense of the overall distribution of information across the items in the current set. Second, users are always presented with a range of relevant attribute values for further narrowing the dataset. Third, in this range users can serendipitously discover relationships among items and attribute values that would otherwise be difficult to discern.

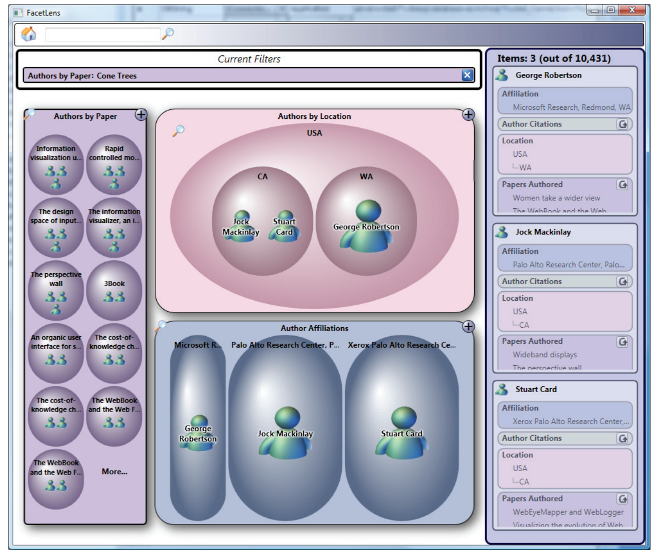

Figure 5. The metadata is shown for the three authors of the Cone Trees paper as a result of the pivot operation.

Attribute co-occurrence relationships are a good example of this third feature in FacetLens. When an item can have multiple simultaneous values for the same attribute, it is considered part of a *multi-value* facet. For example, in a publication dataset, a paper may have multiple authors, making the Papers By Author facet a multi-value facet. Because FacetLens continually reallocates the attribute space to display the most common remaining attribute values, the co-authorship relationship emerges naturally from the visualization. When one author is selected as a filter clause from the Papers By Author facet, the facet remains in the Facet area to display the other authors that occur most often among the remaining papers. For example, FacetLens prominently displays the other authors (e.g., Jock Mackinlay, George Robertson, Peter Pirolli) that co-occur most often on papers where Stuart Card is an author in the InfoVis topic area (Figure 1). This dynamic space allocation is a driving principle in FacetLens and is enforced across all facet types.

Attribute Value Search: Scaling to Large Datasets

A key operation in faceted browsing is to narrow the scope by applying filters on attribute values of the items. Most faceted browsers assume a relatively small number of attribute values. When all the attribute values cannot fit in the allocated screen space, they often provide either a link to a simple list that users have to linearly scan or a paging mechanism to go through successive sets of attribute values.

This becomes difficult when there are thousands of attribute values because of the unpredictable and potentially large number of paging operations and visual searches it entails. For example, with a publication dataset containing thousands of authors, it is difficult or tedious to find an attribute representing a specific author if it does not happen to be presented in the initial view.

To address these issues, in addition to a full-text search box (Figure 1a), FacetLens provides an auto-complete box to allow users to perform a text search for a particular bubble within a facet, scoped only to that facet's attribute values. The magnifier icon, available at the top left corner of each facet (green arrow in Figure 1), reveals a window with a keyword text box and a result list. As users type each key, the list is updated with all partial matches among the attribute values (Figure 6). When users select an attribute value from this window, FacetLens finds the bubble within the facet and selects it as a new filter clause.

Implementation

FacetLens is a standalone Windows application written in C# using Windows Presentation Foundation (WPF) [16], a graphics API for building rich Windows client user experiences. FacetLens sends dynamically generated queries to a Structured Query Language (SQL) database to produce distributions of attribute values and item counts across the current filtered dataset. FacetLens was designed to support multiple datasets without re-authoring. We have used FacetLens on several different datasets, and each required only a few hours of database set up.

FINDING INSIGHTS: EVALUATING EXPERT USE

In order to evaluate the potential of FacetLens in allowing users in their sensemaking tasks, we applied FacetLens to 23 years (1982-2004) of CHI publication data acquired from the ACM. This includes not only full papers but also short papers, demos, and videos, resulting in about 6,000 authors and 4,000 accepted submissions. Unfortunately, the submission types were not demarcated in this dataset.

While we identified some problems with the dataset, we did not make any further effort to correct it since our focus was on the visualization. For example, we did not systematically clean up duplicate author names (e.g., Stuart Card in addition to S.K. Card) or attempt to track down missing or duplicate data. Many of the insights described below could be influenced by this noise in the data.

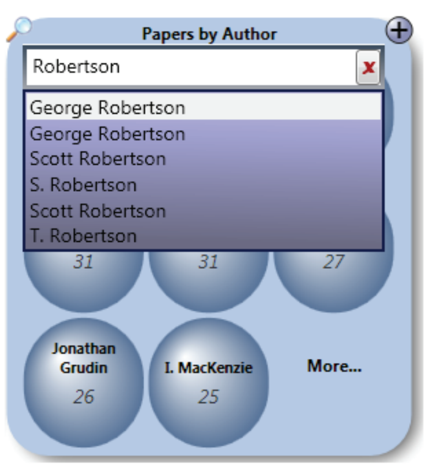

Figure 6. Auto-complete list shows a result list of attribute values containing the keyword the user typed.

One of the authors then spent about an hour or so using FacetLens to explore the dataset in an effort to find interesting trends and relationships. While the author was familiar with usage of the system, he was not quite as familiar with the various datasets. In this section we document both the anecdotal usage patterns, but also the particular insights that the author was able to find by using the system. It is important to note that while some of these insights are found through a targeted search, most of them were found serendipitously by using the system.

FacetLens begins with an overview of the dataset, meaning that we get an interesting summary even before the first user interaction. The number of accepted submissions (not just papers, but all content from 1982 to 2004) in CHI drastically got bigger in 1993 and since then stayed relatively stable (+/-10%) except for 1997, where the data suggests it fell to 77. This anomaly led us to double check numbers and find that data for that year was incomplete. We also see that most authors have USA affiliations. Within organizations, Carnegie Mellon University has published the most (115 accepted) followed by MIT Media Lab (98 accepted) and Georgia Tech (89 accepted). Among individual authors, Brad Myers has published the most (37 accepted). Hiroshi Ishii, who also published a lot (31 accepted), is cited the most (149 times) by CHI papers. Tangible bits, one of his papers, contributes the most to this (92 times). Among paper topics (provided in the dataset), Lab Reports, Applications, Web has the most papers (618 papers). The largest number of authors for a single paper is 15.

After this basic overview, we can explore further by applying filters. To see what has happened with Brad Myers we click on Brad Myers in the Papers By Author facet. FacetLens shows that Brad, for his 37 accepted submissions, has 50 co-authors. He published the most (nine accepted) in the End User Programming topic area. One of his most frequent co-authors is James Landay, with whom he published three papers. We can explore those three papers by moving the mouse over the paper miniature icons. From the Papers by Year facet, we can see that Brad published his first paper in 1985, published the most (six accepted) in 1994, and published at least one paper each year for 15 years.

One of the other authors who had greater familiarity with the CHI conference found it surprising that Brad had not published more than three papers with any given co-author. We used FacetLens to explore this more deeply. First, we used the auto-complete box to search for Brad Myers, and quickly learned that Brad was represented in at least four different ways in the data. This was both because of different versions of his name (e.g., "B. Myers" and "Brad Myers") but also because of an artifact in the underlying ACM dataset that sometimes assigned multiple unique identifiers for the same name. In fact, we found one more instance of James Landay as a co-author, moving the number to four, and this is indeed the case for this dataset.

Turning from authors to topics, we explore the InfoVis topic area by starting over and selecting InfoVis from the Topic facet. By simply clicking on Stuart Card in the Papers By Author facet we can see that Stuart Card published the most (19 accepted) and had 20 co-authors in the InfoVis topic area. He first published a paper in 1987, published the most (four accepted) in 1991, and published at least one paper for 11 years in the InfoVis topic. The co-author who published the most (seven papers) with Stuart Card in the InfoVis topic is Jock Mackinlay. By further clicking on Jock Mackinlay for the Author facet, we can learn that Stuart Card, Jock Mackinlay, and George Roberson wrote four papers together without any other co-authors.

To further evaluate the potential of FacetLens in sensemaking and trend-spotting tasks, and to substantiate the claim that our tool's interface was adaptable to widely disparate datasets, we also applied FacetLens to a database containing 32 years of OECD grant data (1974-2005). Based on obvious attributes of the grant items, the initial view offers facets such as grant donor, grant recipient, year of grant, and grant purpose. At a glance, it is clear that the United States is responsible for donating the highest number, with 98,404 grants out of more than 900,000 total. However, as filters are applied, various other breakdowns reveal themselves. For example, when the grant purpose is restricted to "Culture and Recreation," France becomes the most numerous donor, with more than three times as many grants as the second-place donor, Japan (Figure 7). However, dragging each of these donor bubbles onto the Year of Grant facet reveals in a trend flyout that Japan dramatically increased these grants starting in 2004 and may soon overtake France if the current trend continues. Turning to the grant recipients, we select the Europe region and find that the vast majority of grants to Europe over the past 32 years have gone to countries representing the former Yugoslavia. One interesting exception is Gibraltar, which has received a smattering of development grants (41) over the years, nearly all of them with the United Kingdom as donor.

USABILITY STUDY: EVALUATING NOVICE USE

In the previous section we provided evidence that an expert user could utilize FacetLens to gain understanding of the CHI dataset. In this section we describe our efforts to test whether or not novice users could learn to use the system with the same efficacy. Specifically, we ran a formative usability study using the same dataset to identify areas of confusion and opportunities for improvement.

We recruited five researchers and one developer (two females) from within our research organization. All of the researchers had expert understanding of academic publication and were familiar with CHI, though the developer was slightly less familiar. The average age of our participants was about 33, ranging from 23 to 39 years old. None of the participants had interacted with FacetLens before the study. We ran participant one at a time with each session lasting about 30 minutes. FacetLens was run at a resolution of 1200×1000 on a 24″ 1920×1200 Dell display.

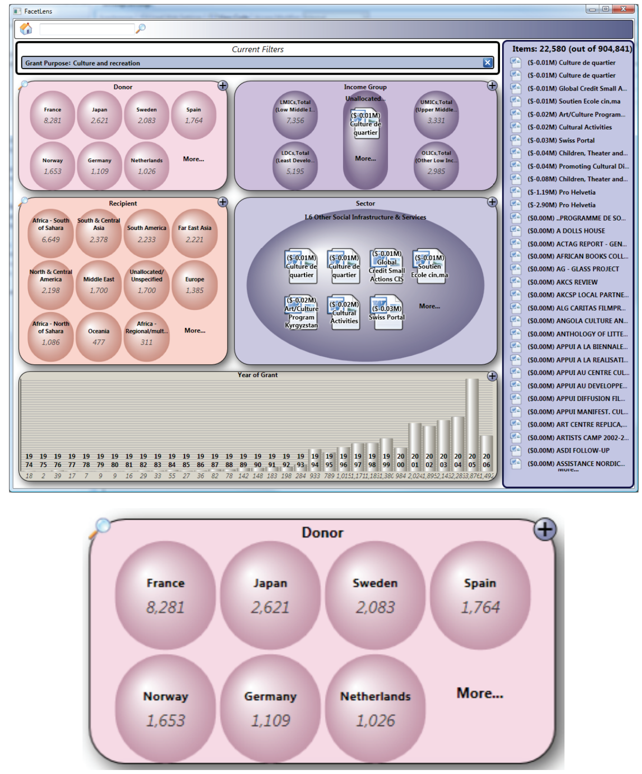

Figure 7. France is the top donor for the "Culture and Recreation" grant (enlarged view of Donor facet on bottom).

We chose tasks that began simply and then gradually became more complex (i.e., the task semantics became more complex, and/or multiple filters and interactions were required). For example, simple tasks include "Which author publishes the most frequently in the topic of CSCW?" and "What are the affiliations of the two authors from China?" More complex ones include "Compare the publication trends for Lab Reports by year and CSCW papers by year. How do they differ?" and "Describe three facts about Randy Pausch." Some tasks consisted of two or three subtasks. For example, 1) Find the authors of the paper "The Perspective Wall." Which states did they live in? and 2) Pivot to the papers authored by Jock Mackinlay. Which paper did his papers cite the most?

We gave participants a brief scripted tutorial at the beginning of the experiment, after which we handed them the task list and asked them to "think out loud" as they performed each task. We took observation notes and noted the time for each task (rounded to the nearest minute). After all tasks were completed, participants filled out a user satisfaction questionnaire, and provided any feedback they had.

Results

Tasks were completed relatively quickly. Most tasks were accomplished in less than one minute, but a few (in particular, tasks 4, 6a and 9) took a little longer, on average. This was expected, as we purposefully built the task list to grow in difficulty due to the increasing complexity of the queries required to find the answers.

Usability Issues

The main goal of running a formative usability study is to identify issues with the user interface so that we can iteratively redesign those problem areas.

One of the major issues pointed at confusion over the symmetric nature of the relationships of some facets to each other. This was further compounded by the fact that the terminology used in publication data inherently makes it difficult to parse directionality. For example, if the user filtered on a particular author, say Brad Myers, the Author Citations facet represented all authors that were cited within any of Brad's papers, Paper Citations represented all papers cited by Brad's papers, and References represented papers that referenced any of Brad's papers. This is confusing and we are considering methods of exposing the directionality of such item-to-item relationships.

Another issue had to do with understanding that, once a particular author was selected from the Papers by Author facet, the semantics of the facet now changed subtly to indicate co-authors. Because of the complexity introduced by centering the system around both papers and authors, users had to stop and think about what it meant for a set of authors to remain when they had already filtered by one.

We were initially concerned that the drag and drop interacttion was less intuitive than clicking, but participants seemed to get this after a single demonstration, even to the point of quickly generalizing that they could do this for multiple attributes at a time.

Search within the facets themselves worked very well for participants. They especially liked the auto-complete feature that let them know what was available as a search term. Tooltips did not show up on hover for the facet itself, which sometimes caused problems in understanding what a facet title meant in a particular context. Also, in the item list, the user had to click on a document or author's icon, not the title text next to the icon, in order to get details on that item. Overall, however, the user interface was used quickly and efficiently for a first time evaluation.

User Satisfaction

As mentioned, users filled out satisfaction surveys after completing all tasks. The average user satisfaction scores were quite high for such a novel user interface with so little training (the scale was 1 to 7, with Disagree, Agree as the anchors of the scale). Users liked using this system (avg. score 6.1) and the system seems to be easy to learn (avg. 5.7). Furthermore, it seems to be easy to discover trends using the "Papers by Year" facet (avg. 6.1). Because the pivot operation was not fully understandable, there is still room for improvement (avg. 5.4). The most notable average low ratings centered on the organization of the interface being clear (avg. 4.9), and this is likely due to the confusion around some of the facet names and their symmetrical nature, as noted above.

Error recovery also seemed to be an issue (avg. 4.7), so we plan on focusing on better ways to "undo" or roll back changes. It would appear that the animations and transition speeds were not unusable and did not cause direct trouble, but could be improved in the future as well. We intend to improve all these issues in the next iteration of our design.

DISCUSSION AND FUTURE WORK

As stated above, FacetLens was used efficiently by participants and rated fairly highly for its first round evaluation. There are clearly areas of improvement: adding clarity to the facet relationships and directionality; using color; and trying different animation and transition timing and types. Many of these usability issues are easy fixes, though some require more thought. In future work, we intend to iterate on the system and deploy to a wider set of end users.

Through this work, we have also identified other features that could improve the FacetLens user experience. For example, there is no easy way to find all papers written by all authors in Japan since users can perform a pivot operation only from an individual item. We also would like to examine ways to scale to even larger datasets such as ACM Digital Library (DL) with many other kinds of metadata. One of the challenges we expect is that we will have many more facets. In addition to needing a way to handle a larger number of facets, we need a better way to identify interesting and meaningful facets. We also need to provide a way to configure and save a starting point, or "home" view. For example, most people will probably be interested in only part of the ACM DL, meaning that they would want to apply several filters by default at start up. More generally, it would be useful to allow users to save the current state, consisting of a set of filters, for later use.

In FacetLens, we tried to break the firm division between attributes and items by allowing items to be attributes of other items. This means FacetLens is not designed to be centered on a single item "type." Depending on what filters users apply, it shows a view for a specific item type. For example, if users apply a Paper by Author filter, FacetLens becomes a paper view. If users apply an Authors by Location filter, it becomes an author view. User comments from the usability study, at least for the CHI publication dataset, lead us to suspect that it might be easier to explore the dataset if we provide users with a purely paper-centric visualization. Even though Authors is a valid item type, it may be that presenting items with disjoint sets of metadata side by side is simply too complex for non-expert users to use effectively. We are planning to investigate this further by changing the facet definitions in the database to compare the current data model with a more paper-centric model.

CONCLUSION

In this paper, we have described FacetLens, an interactive visualization system that supports exploration and sensemaking within faceted datasets by exposing trends and relationships. FacetLens introduces several new interactions, features, and facet types. Linear facets enable users to easily identify trends, and simple drag-and-drop operations combined with facet flyout capabilities allow users to easily compare several trends at once. Furthermore, FacetLens offers pivot operations to allow users to navigate faceted data using the relationships between items.

To show the utility of the system, we reported insights gained while experts used the system to explore the CHI publication repository and an OECD database of grant data. We also presented results from a formative user study conducted to identify usability issues. FacetLens was used easily by our participants in exploring conference proceedings content, and overall, participants liked it. Finally we discussed potential future work for improving FacetLens and scaling it to larger repositories.

ACKNOWLEDGEMENTS

The authors would like to thank the user study participants.

REFERENCES

1. Ahlberg, C. Spotfire: an information exploration environment. *ACM SIGMOD Record*, 25, 4 (1996), 25-29.
2. Bungee View, http://cityscape.inf.cs.cmu.edu/bungee
3. Gapminder, http://www.gapminder.org
4. Havre, S., Elizabeth, H., Whitney, P., and Nowell, L. ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Trans. Visualization and Computer Graphics*, 8, 1 (2002), 9-20.
5. Kang, H., Plaisant, C., Lee, B., and Bederson, B.B. NetLens: Iterative Exploration of Content-Actor Network Data. *Information Visualization Special Issue on Visual Analytics*, 6 (2007), 18-31.
6. Klein, G., Moon, B., and Hoffman, R.F. Making sense of sensemaking I: alternative perspectives. *IEEE Intelligent Systems*, 21, 4 (2006), 70-73.
7. Lee, B., Czerwinski, M., Robertson, G., and Bederson, B.B. Understanding research trends in conferences using PaperLens. *Extended Abstracts CHI 2005*, ACM Press (2005), 1969-1972.
8. Mackinlay, J., Hanrahan, P., and Stolte, C. Show Me: Automatic presentation for visual analysis. *IEEE Trans. Visualization and Computer Graphics*, 13, 6 (2007), 1137-1144.
9. Microsoft Office Excel, http://office.microsoft.com/en-us/excel
10. Ranganathan, R. Colon Classification: Basic Classification. Sarada Ranganathan Endowment for Library Science (1991).
11. Robertson, G., Fernandez, R., Fisher, D., Lee, B., and Stasko, J. Effectiveness of animation in trend visualization, *IEEE Trans. Visualization and Computer Graphics*, 14, 6 (2008), 1325-1332.
12. Smith, G., Czerwinski, M., Meyers, B., Robbins, D., Robertson, G., and Tan, D. FacetMap: A scalable search and browse visualization. *IEEE Trans. Visualization and Computer Graphics*, 12, 5 (2006), 797-804.
13. Spenke, M. and Beilken, C. InfoZoom - Analysing formula one racing results with an interactive data mining and visualization tool. *Proc. Data Mining 2000*, (2000), 455-464.
14. Stolte, C., Tang, D., and Hanrahan, P. Polaris: a system for query, analysis, and visualization of multidimensional relational databases. *IEEE Trans. Visualization and Computer Graphics*, 8, 1 (2002), 52-65.
15. Wattenberg, M. Baby names, visualization, and social data analysis. *Proc. InfoVis 2005*, IEEE (2005), 1-7.
16. Windows Presentation Foundation (WPF), http://windowsclient.net
17. Wong, P.C., Hetzler, B., Posse, C., Whiting, M., Havre, S., Cramer, N., Shah, A., Singhal, M., Turner, A., and Thomas, J. IN-SPIRE InfoVis 2004 Contest Entry. *Posters Compendium of InfoVis 2004*.
18. Yee, P., Swearingen, K., Li, K., and Hearst, M. Faceted metadata for image search and browsing. *Proc. CHI 2003*, ACM Press (2003), 401-408.
19. Zhang, J. and Marchionini, G. Evaluation and evolution of a browse and search interface: Relation Browser++. *Proc. Digital Government Research 2005*, (2005), 179-188.