Faceted Exploratory Search Using the Relation Browser

Robert G. Capra and Gary Marchionini
School of Information and Library Science
University of North Carolina at Chapel Hill
rcapra3@unc.edu, march@ils.unc.edu

ABSTRACT

The Relation Browser (RB) is a tool developed by the Interaction Design Lab at the University of North Carolina at Chapel Hill for understanding relationships between items in a collection and for exploring an information space (e.g., a set of documents or web pages). It is implemented as a Java Applet that can be embedded in a web page. The Relation Browser provides a dynamic user interface that allow users to explore a data set through the use of faceted browsing and keyword search. The current Relation Browser, RB07, is available at: http://idl89.ils.unc.edu/rb07

1. Relation Browser 07

The Relation Browser has a long history and has been through a number of significant revisions [5][7]. At JCDL 2007 [2], we reported on two studies we conducted to compare three different interface styles (a handcrafted web site, a simple facet interface, and the previous RB++ version of the Relation Browser) for three different task types (simple lookup, complex lookup, and exploratory search) for the U.S. Bureau of Labor Statistics (BLS) web site data. Based on the results of the studies and on expert reviews of the interface, we developed and implemented a set of design changes for the next-generation RB [1][3], dubbed the RB07, while still maintaining a primary goal of providing a tool for exploring data spaces – especially for gaining a better understanding of documents and how they are related to each other.

The Relation Browser has two "hallmark" features. First, all the facet values are displayed "front and center" in the interface, with visual components (bars) that indicate the number of matching documents for each facet value (see Figure 2). Second, the interface supports dynamic queries [6], meaning that as you mouse-over facet values, the display is automatically updated to show a preview of including the moused-over facet value in the current query. This can cause the length of the bars to change as well as the results shown in the results area. These and other features of the RB are designed to support exploratory search [4].

Additional features of the RB07 are described below (numbers correspond to the numbered circles in Figure 2).

1. Multiple facet views – The RB07 supports multiple, pluggable facet views. Users can switch between views using tab controls. The "Facet List" view presents the facets and facet values as lists of TriBars (described below) that indicate the number of matching documents. In addition, there is a "Facet Cloud" view (Figure 4) that displays the facet values in the style of a tag cloud, using font size to indicate the number of matching documents. Both the Facet List and Facet Cloud dynamically update on mouse-overs.

2. Static facet list – An additional listing of the facets was added down the left side of the screen to provide a constant view of the facets that does not dynamically update (except to change the color of facets in the current query). This is especially useful when working with the facet cloud since the position of the facet values in the cloud dynamically update as the mouse is moved within the cloud.

3. Multiple result views – Results of our RB++ study revealed that many users were accustomed to search engine style interfaces that display results in linear list. However, the RB++ compactly displayed results in a grid. The new RB07 allows users to select either the grid view of the prior RB++, or a list view that is similar to search engine result displays.

4. Current query display and control – In the new RB07, the current query is displayed near the top of the screen and acts similarly to breadcrumb trails that many web sites use – it shows a history of the facets and search terms entered and provides buttons to remove individual facets or terms. This allows users to quickly remove items from an over-constrained query.

5. and 6. Full-text search and search within the results – Observations and results from our RB++ study indicated that many users wanted to employ a "keyword search first" strategy to information seeking tasks. The RB++ did not support full-text search and encouraged a "facets first" approach to exploratory search. In the new RB07, we have sought to support the use and mixing of both strategies by including full-text search and search within the results, while maintaining a prominent display of the facets and document space.

2. TriBar Displays

The list view of the Relation Browser uses a custom user interface element called a TriBar (see Figure 1) that visually displays three pieces of information about each facet value.

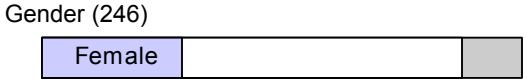

Figure 1. An example TriBar display

The total number of documents that match this facet is shown in parentheses next to the name of the facet (e.g. "Gender (246)"). The full width of the TriBar (the right edge of the grey area) represents the total number of items *in the collection* that have this *facet* (e.g. 246 items have the Gender facet). Second, the right edge of the white area indicates the total number of items *in the collection* that have this *facet value* (e.g. About 90% of the items that have a Gender facet have the value Female). Note that items may have more than one facet value for each facet, as is the case with the Gender facet shown in Figure 2 (many documents discuss both males and females in the same document). Third, the right edge of the purple area on the left side indicates the total number of items *in the current query* that have this *facet value*.

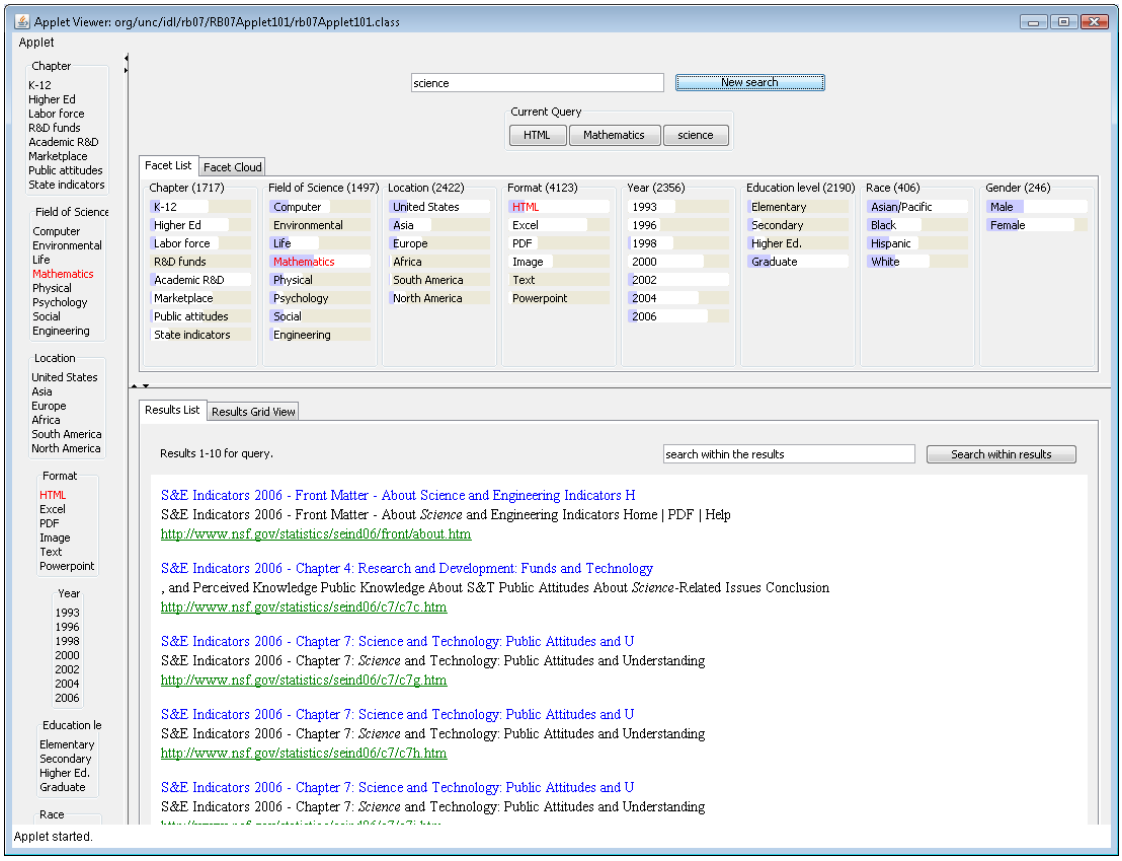

Figure 2. Screenshot of the Relation Browser 07

3. Architecture

The RB07 is implemented as a Java Applet that communicates with an Apache web server and Apache SOLR search engine as shown in Figure 3.

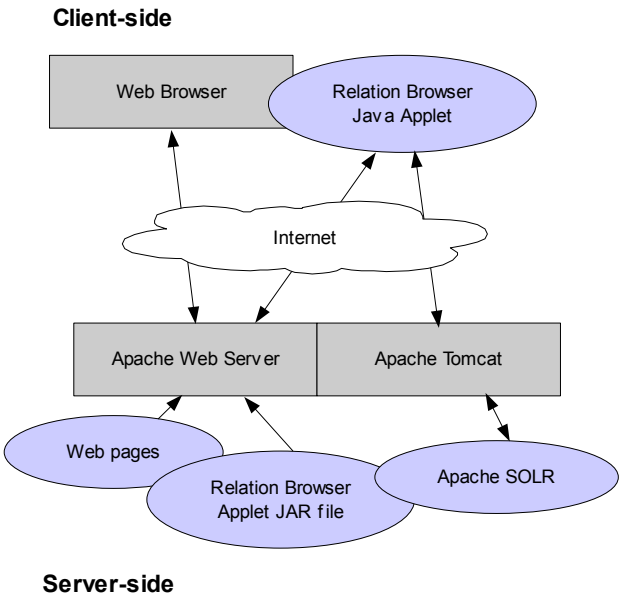

Figure 3. Relation Browser Architecture

When the applet is started in the client browser, it connects to the server and loads the facet data so that it can calculate and display the facet counts in the TriBars. Facet data is stored and transmitted as a compressed bit set to minimize the transfer time from the server to the client. All computation and updating of the facet counts is done on the client so that the dynamic display can be updated smoothly. Keyword search is handled using an Apache SOLR server (based on the Lucene search engine). When a keyword search is entered, the query is sent to the SOLR server and the results are merged (on the client side) with the existing facets in the current query. Keyword search results are cached so that subsequent query manipulations do not require a re-query of the SOLR server.

4. Configuration

Each Relation Browser instance is configured using two types of XML configuration files. The first type of configuration file is an instance file that describes the facets, facet values, and other properties of the Relation Browser. There is only one instance file. An example is shown below.

```
<rbinstance>
  <resultfields>
    <resultfield>docid</resultfield>
    <resultfield>Title</resultfield>
    <resultfield>URL</resultfield>
    <resultfield>Description</resultfield>
  </resultfields>
  <fulltextfields>
    <fulltextfield>fulltext</fulltextfield>
```

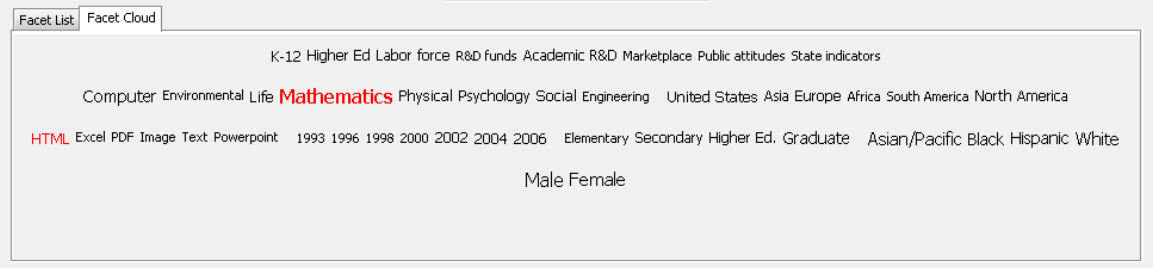

Figure 4. Screenshot of the Facet Cloud View of the Relation Browser 07

```
  </fulltextfields>
 <facets>
   <facet>
     <facetname>Chapter</facetname>
       <facetvalues>
        <facetvalue>K-12</facetvalue>
        <facetvalue>Higher Ed</facetvalue>
        <facetvalue>Labor force</facetvalue>
        <facetvalue>R&amp;D funds</facetvalue>
        <facetvalue>Academic R&amp;D</facetvalue>
        <facetvalue>Marketplace</facetvalue>
        <facetvalue>Public attitudes</facetvalue>
        <facetvalue>State indicators</facetvalue>
       </facetvalues>
   </facet>
   . . .
```

The second type of configuration file are the document description files. There is one document description file for each document in the collection. These files specify what facets, fields, and values apply to each document. An example is shown below.

```
<doc>
   <field name="docid">125</field>
   <field name="URL">
     http://www.nsf.gov/statistics/seind06/c1/c1s2.htm
   </field>
    <field name="Description"> S&amp;E Indicators 2006
- Chapter 1: Elementary and Secondary Education -
Student C</field>
   <field name="Year">2006</field>
   <field name="Chapter">K-12</field>
   <field name="Field of Science">Mathematics</field>
   <field name="Location">United States</field>
   <field name="Format">HTML</field>
   <field name="Year">2006</field>
   <field name="Education level"> Elementary</field>
   <field name="Education level">Secondary</field>
   <field name="Education level">Higher Ed.</field>
   <field name="Education level">Graduate</field>
    <field name="fulltext"> S&amp;E Indicators 2006 -
Chapter 1: Elementary and Secondary Education -
Student Coursetaking in Mathematics and Science
Home | PDF | Help | Contact Us Search Table of
Contents Overview Chapter 1 Highlights
   . . .
```

5. Future Work

We plan to evaluate the features of the new RB07 using a method similar to the one we used in our JCDL 2007 paper [2] to evaluate the previous RB version. That is, we plan to compare it to a "vanilla" faceted interface, an hand-crafted web site, and a basic search engine interface. We are also interested in evaluating the usefulness and usability of the facet cloud and related visualizations of the facet space.

6. Acknowledgments

This work was supported by a contract with the NSF Division of Science Resources Statistics, Information and Technology Services Program.

7. References

1. Capra, R., and Marchionini, G. (2008). The Relation Browser Tool for Faceted Exploratory Search. Proceedings of the 2008 Conference on Digital Libraries (JCDL '08), Pittsburg, Pennsylvania, June 16 - 20, 2008.

2. Capra, R., Marchionini, G., Oh, J. S., Stutzman, F., and Zhang, Y. (2007). Effects of structure and interaction style on distinct search tasks. In Proceedings of the 2007 Conference on Digital Libraries (JCDL '07), Vancouver, BC, Canada, June 18 - 23, 2007.

3. Capra, R., and Marchionini, G. (2007). Visualizing Science and Engineering Indicators: Transitioning from Print to a Hybrid World. Technical Report TR-2007-07, School of Information and Library Science, University of North Carolina, Chapel Hill, NC.

4. Marchionini, G. (2006) Exploratory search: From finding to understanding. *CACM*, 49(4), 41-46.

5. Marchionini, G. and Brunk, B. (2003) Towards a general relation browser: A GUI for information architects. *Journal of Digital Information, 4*(1).

6. Shneiderman, B. (1994). Dynamic Queries for Visual Information Seeking, *IEEE Software*, 11(6): 70-77.

7. Zhang, J. and Marchionini, G. (2005). Evaluation and evolution of a browse and search interface: relation browser. In *Proceedings of the 2005 National Conference on Digital Government Research* (Atlanta, Georgia, May 15-18, 2005). ACM International Conference Proceeding Series, vol 89. Digital Government Research Center, 179-188.