Backward Highlighting: Enhancing Faceted Search

Max L. Wilson, Paul André, m.c. schraefel
School of Electronics and Computer Science
University of Southampton
Southampton, UK
{mlw05r,pa2,mc}@ecs.soton.ac.uk

ABSTRACT
Directional faceted browsers, like the popular columnar browser iTunes, let a person pick an instance from any column-facet to start their search for music. The expected effect is that any columns to the right are filtered. In keeping with this directional filtering from left to right, however, the unexpected effect is that the columns to the left of the click provide no information about the possible associations to the selected item. In iTunes, this means that any selection in the Album column on the right returns no information about either the Artists (immediate left) or Genres (leftmost) associated with the chosen album.

Backward Highlighting is our solution to this problem, which allows users to see and utilize, during search, associations in columns to the left of a selection in a directional column browser like iTunes. Unlike other possible solutions, this technique allows such browsers to keep direction in their filtering, and so provides users with the best of both directional and non-directional styles. Providing the technique is not cheap, however, as it produces significantly more queries over the data. As well as describing the detail of Backward Highlighting, this paper presents the results of a formative user study into the benefits the technique provides to users. System designers can then determine whether the technique should be used given both the system costs and the significant benefits shown for information discovery and subsequent retention in memory.

ACM Classification: H5.2. User Interfaces: Graphical user interfaces. H.3.3. Information Search and Retrieval: Information Filtering.

General terms: Design, Human Factors.

Keywords: Faceted, search, columns, highlights, context.

INTRODUCTION
Visualising and manipulating multi-dimensional information spaces is a known hard problem. Finding solutions is becoming increasingly important as the information available to us is growing exponentially, and creating novel and effective ways to explore this data is critical. An example approach to visualising and interacting with these kinds of information spaces is faceted browsing, which has been used by many organisations including IBM, Boeing, Wal-Mart, eBay (Yahoo) and Apple. In this paper we present a lightweight technique for improving the interactivity, learnability and explorability of faceted representations for large information spaces.

There are two main approaches to faceted browsing: non-directional faceted browsers and directional column-faceted browsers, both of which are discussed further in related work. Despite the additional metadata that can be conveyed by using the directional column-faceted approach, like iTunes[1] and mSpace [8], if a user makes their first selection in a middle column-facet, and the browser only filters columns to the right, then there will be no associated information indicated in all the columns to the left of the selected item. Associated information will only be made visible in the filtered lists in the columns to the right of their selection. This is shown clearly in Figure 1, where an Artist is selected in the middle column of iTunes, and related Album's are shown to the right, but no associated Genres are shown on the left. Consequently, the Genre column is excluded from contributing to the search, unless the user already knows about the Genres of the selected Artist.

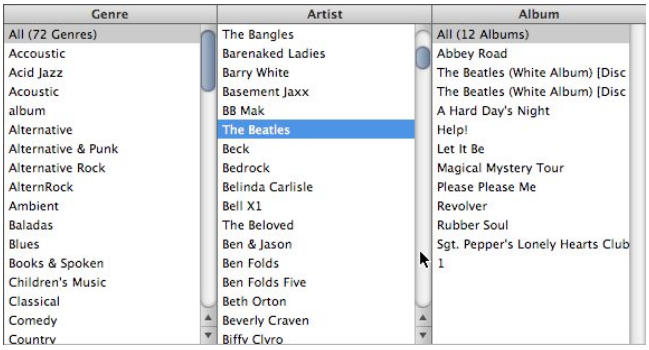

Figure 1: iTunes showing The Beatles selected, with their Albums showing in the Album column to the right, and no associations shown in the Genre column to the left.

We propose a technique for mitigating this identified flaw of the directional approach, by revealing associations in columns to the left of a selection, which we call Backward Highlighting (BH). In the example above, BH would reveal the missing associations in the Genre column with highlights so that the potentially selectable relationships to the

[1] http://www.apple.com/itunes/

left, can then be used to further refine the search for a desired album or song. The BH technique, therefore, converts previously passive and uninformative columns into interactive visualisations of rich metadata associations.

Although a fairly intuitive idea, it has not been included in any directional column-faceted browsers like iTunes. Possible reasons for the non-appearance of such a tool could be the known technological challenges, investigated by Smith et al. [9]. Each additional column being backward highlighting, adds to the load of queries over the dataset. Alternatively, the potential variables in design and unmeasured benefits to users are unknown, but investigated below.

In the remainder of the paper, we first describe related work, including more detail on the different approaches to the big business of faceted search. We then describe the BH technique in more detail and review a formative evaluation of its benefits, by comparing two possible implementations. We conclude with design recommendations and by summarizing the benefits of using the directional column-faceted approach, where including BH provides users with the best of both faceted approaches.

RELATED WORK

Increasingly, one popular interactive alternative to keyword search has been faceted browsing [4], and has been shown to improve over other techniques such as browsing basic hierarchical categories [2] and clustering [5]. Further, Wal-Mart, IBM, and Boeing are just three examples of multi-million dollar organizations that are using Endeca's faceted search[2]. eBay[3], owned by Yahoo, is also using faceted search to help potential bidders find things they want to buy. Further, as mentioned above, Apple is using faceted search within their software; iTunes has over 200 million users [7]. Even Google, whose keyword search box has become so familiar on the web, provide faceted search along with their product searches[4].

The premise of faceted browsing, which is well described by Capra et al. [1], is to provide the whole range of metadata around a document collection as optional constraints during search. For example, Classical Music may have facets such as Song Titles, Composers, Periods, Instruments, and Arrangements, where songs have one or more values in each facet. While looking for classical music, users may want to narrow their results by items in any or all of these facets, depending on their existing knowledge. Figure 2 shows an mSpace [8], as an example of a faceted browser that can be used to explore classical music. Three of the facets are presented in a row across the top of the interface in a similar way to iTunes.

Part of the popularity of faceted search is that it provides an alternative or complementary search approach to the familiar keyword search provided by services like Google. Providing facets or categories instead allows users to perform more exploratory forms of search [10], such as browsing, exploring, and evaluating options. Further, however, users have been seen to also use facets to construct very rich queries over a dataset in preference to boolean queries or using advanced search forms [11].

Non-directional and Directional Faceted Browsing

Given all these benefits provided by faceted search, there are some significant variations in the way that they are implemented. Most notable is the difference between non-directional faceted searching, such as Flamenco [3] and Exhibit [6] and directional column-faceted browsers like iTunes and mSpace [8]. In the non-directional design of faceted browsers, a user makes a selection in one facet and all of the facets are filtered to show only the items that are relevant to the item selected, including the facet with the selection. This design, while effective for quickly narrowing down a search, presents fewer persistently available facts as it removes all the information from the screen that is unrelated to the combined selections. In directional column-faceted search there is a specific order of facets (usually in columns) and the filtering usually goes from left to right; this is true of both mSpace and iTunes. That is a selection in a left column filters the columns to the right.

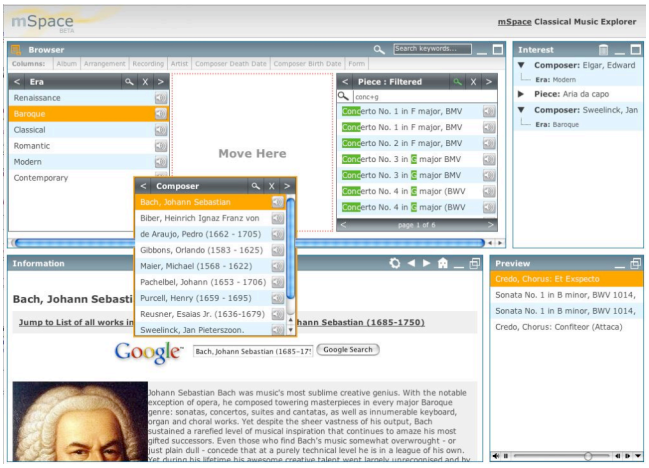

Figure 2: The mSpace directional column-faceted browser, showing three column-facets on classical music: Era, Composer and Piece Title.

Let us consider a faceted system that allows users to browse classical music, where each click filters every facet. Facets might include the Eras, Composers, Arrangements, and Pieces of music. Selecting a Composer in a non-directional faceted system would remove all of the Eras, Composers, Arrangements, and Pieces not related to that Composer. A further selection of a type of Arrangement, removes all of the Eras, Composers, Arrangements, and Pieces unrelated to the selected Arrangement by the previously selected Composer. The user is now unable to see all of the Arrangements that the composer produced, as all of the other Arrangements, and indeed the other Composers, are not related to the selected Arrangement by the selected Composer. In directional column-faceted browsers selecting a Composer would filter the Arrangement and Piece columns, but leave all the composers in view. Then select-

[2] http://endeca.com/customers/index.html

[3] http://express.ebay.com

[4] http://google.co.uk/products?q=oven&btnG=Search+Products

ing one of the remaining Arrangements would filter the Piece column only, leaving all the Arrangements used by the Composer still in view. There is no benefit in removing the Composer's contemporaries and the other types of arrangements, unless screen real-estate is an issue. In the remainder of this paper, we call this type of information that is usually lost in non-directional faceted search, but kept in directed faceted search, 'Added Facts'.

BACKWARD HIGHLIGHTING

Although directional column-faceted browsers can convey additional information to users, the iTunes example in the introduction shows that some columns can become passive and uninformative. The intuitive proposition is to highlight associations in columns to the left of a selection to indicate the possible paths a user could have taken to get to where they chose to start. This concept turns passive and uninformative columns into interactive visualizations of rich associations that could be further used to refine a search. As this highlighting happens backwards, against the left-to-right flow of filters, we call it Backward Highlighting.

Figure 3: The Cello Arrangement is selected and the Piece column shows all the Pieces with a Cello Arrangement. Backward Highlighting shows the Eras and Composers, such as Schumann and Vivaldi, associated with Cello music. This shows that the Cello music on the right could come from one of five Eras and one of twelve composers.

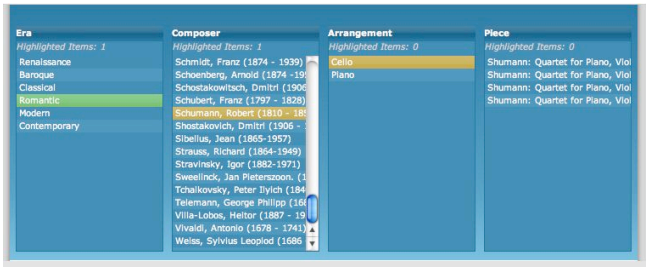

Figure 4: Continuing from Figure 3, one of the associated Composers, Robert Schumann, who was previously highlighted as related to Cello music, has now been selected. The Romantic Era has been Backward Highlighted as it is now the only Era relating to Schumann's Cello music. The user has refined their song results in a way that was not possible without Backward Highlighting.

Let us again consider the classical music example, as shown in Figure 3 and Figure 4. Where we have the Era, Composer, Arrangement, and Piece facets, should a user select a certain Arrangement, the potentially multiple associated Composers and Eras would simply highlight and the associated Pieces would remain in the final column (Figure 3). The user can now see all of the composers who used the selected type of arrangement and could select any one of them to further refine their search. The ability to make a further selection to the left of the initial selection was previously the cost of directional column-facet browsers but was possible in non-directional faceted search. Now able to do this in directional browsers, selecting a Composer would leave only the Era of the selected Composer's Arrangements highlighted (Figure 4). The Piece column, as always, would show only the Pieces of that Arrangement by the selected Composer. The visible Added Facts, which would not be seen in non-directional faceted search, are: 1) the Eras where the Composer did not compose any pieces in the selected Arrangement, and 2) all of the Arrangements used by the Composer, including the selected one.

The overall effect of adding the BH technique is that the user gets the benefit of both non-directional and directed faceted browsing, increasing flexibility of interaction but maintaining some sense of structure for the user. In terms of increasing domain knowledge, however, the expected effect was that the user will see further associated information, along with their selections and filtered lists, and may result in increased discovery and retention of additional facts. In terms of their future search behaviour, it was expected that the BH technique allows users to convert their increased domain knowledge and use previously highlighted items as filters. By doing so, they would not have to scroll through long lists, such as the Composers column, but filter them by the Era the composed in.

EVALUATION

To better understand the benefits of BH, a study of 18 participants (12 male), was conducted on three interfaces with the BH technique, and its implementation, as the independent variable: no BH, BH, and a version where the highlights were grouped at the top of each column. So that any affect could be directly attributed to the columns and highlights, no further content or search results were displayed on the interface. Three datasets were used, the exposure to each interface and dataset was rotated.

Given that BH highlights metadata related to selections, three hypotheses were identified: H1) that users would be able to discover more facts with some form of highlighting; H2) that users would remember more facts; and H3) that users would be able to use these remembered facts to improve their search behaviour.

To evaluate these hypotheses, there were three phases of the 90 minute participation period: finding and writing facts about an item in one of the middle two facets, re-finding facts with the interface, and remembering and writing down facts that had been learned.

Results

We present the results of the study under four headings to provide a range of understanding about the BH technique: measured discovery of facts, measured retention of facts, affect on behaviour, understandability, and preference.

Measured Discover of Facts. As hypothesized (H1 accepted), there was a statistically proven increase in the number of left-of-selection facts discovered by participants in the experimental conditions over the control condition (ANOVA, $F=14.97$, $p<0.0001$) and there was no significant effect on right-of-selection facts. A t-test revealed no significant difference between the grouped and ungrouped conditions.

Deeper analysis by dataset, however, revealed that significant benefits (t-test, $t=2.7269$, $p<0.05$) were found when the left-of-selection columns were longer. Deeper analysis into what we termed 'Added Facts', those provided by the directional faceted approach, revealed that a significant number were found with the non-grouped experimental condition (ANOVA, $F=3.869$, $p<0.05$). This suggests that highlights should not be grouped to encourage discovery.

Measured Retention of Facts. As hypothesized (H2 accepted), participants were able to recall significantly more facts from the highlights (ANOVA, $F=5.292$, $p<0.01$), although no significant difference was found between the grouped and ungrouped conditions.

Changed Behaviour Patterns. Contrary to our hypothesis (H3 rejected), during refinding tasks, the measured addition of facts did not have a consistent affect on behaviour. More work is required to understand what level of experience or increased knowledge has an affect on behaviour.

Understandability. Although subjective ratings of difficulty showed a significant improvement for the experimental conditions (ANOVA, $F=5.715$, $p<0.005$), there was no significant difference between the two experimental conditions. Further in a longitudinal study of mSpace, which included the ungrouped BH condition, log analysis and comments indicated that users were comfortable with the highlighting behaviour during the first search session [11].

Preference. 16 out of 18 preferred grouped highlights. Further qualitative comments about the colour of highlights from our longitudinal study have indicated that users prefer the highlights to be a fainter version of the colour used for selections. A whole separate colour was found to be confusing to some participants.

DESIGN RECOMMENDATIONS

The results above show that users were able to discover significantly more facts and retain more of them in memory with both grouped and ungrouped highlighting over no backward highlighting. The results also show that grouped was only significantly better than backward highlighting in the context of long lists (greater than 200 items in a parent list). The cost of grouping, however, is that users were seen to explore the associations and Added Facts less. The strongest evidence for grouping highlighted items was from user preference, however qualitative comments suggest that the preference is in the power of the grouped items. Consequently, also considering the conditional and statistical evidence, we suggest that on-demand grouping should be available for users, with ungrouped as the default. Aesthetically, we suggest that colour of highlights should be tied to the colour of selections, so it is clear to the users that highlights do not filter but can be applied through interaction.

CONCLUSIONS

In this paper we have introduced and evaluated a technique we term Backward Highlighting - a contextual cue for directional column-based faceted browsers. By highlighting associations in columns to the left of a selection, the BH technique converts a previously static unassociated list in two ways. First, the highlighted items in lists convey new information that is not seen without the use of Backward Highlighting. Second, the highlighted items can be used to refine an existing set of selections, where a user was previously constrained to column-facets to the right only. Consequently, when Backward Highlighting is included within the directional faceted approach, browsers like iTunes and mSpace have all the advantages of both non-directional and directional approaches to faceted browsing, and none of the known disadvantages. Such results are important for businesses currently using faceted search (Apple, eBay, IBM, Wal-Mart) that strive to make their products more accessible to users.

ACKNOWLEDGMENTS

Thanks to Steven Day for work on implementation, and the participants of the studies.

REFERENCES

1. Capra, R., Marchionini, G., Oh, J.S., Stutzman, F. and Zhang, Y. Effects of structure and interaction style on distinct search tasks. *JCDL07*2007). 442-451.
2. Chimera, R. and Shneiderman, B. An Exploratory Evaluation of Three Interfaces for Browsing Large Hierarchical Tables of Contents. *ACM TOIS*, *12* 4 (1994). 383-406.
3. Elliott, A., Flamenco image browser: using metadata to improve image search during architectural design. *CHI extended abstracts*, (2001). 69-70
4. Hearst, M.A. Next generation web search: setting our sites. *IEEE Data Engineering Bulletin: Special Issue on Next Generation Web Search*, *23* 3 (2000). 38-48.
5. Hearst, M.A. Clustering versus faceted categories for information exploration. *Commun. ACM*, *49* 4 (2006). 59-61.
6. Huynh, D.F., Miller, R. and Karger, D., Exhibit: Lightweight Structured Data Publishing. *WWW07*, (2007).
7. Kahney, L. Apple Gives Video the IPod Touch *Gadgets, Mac*, 2005.
8. schraefel, m.c., Wilson, M.L., Russell, A. and Smith, D.A. mSpace: improving information access to multimedia domains with multimodal exploratory search. *Commun. ACM*, *49* 4 (2006). 47-49.
9. Smith, D.A., Owens, A., schraefel, m.c., Sinclair, P., André, P., Wilson, M.L., Russell, A., Martinez, K. and Lewis, P., Challenges in Supporting Faceted Semantic Browsing of Multimedia Collections. *SAMT07*, (2007).
10. White, R.W., Kules, B., Drucker, S.M. and schraefel, m.c. Introduction. *Commun. ACM*, *49* 4 (2006). 36-39.
11. Wilson, M.L. and schraefel, m.c., A longitudinal study of both faceted and keyword search. *JCDL08*, (2008).