Impact of Visualization Methods on Interaction with Search Results

Jennifer Golbeck, Chang Hu
Human-Computer Interaction Lab
University of Maryland, College Park, MD
jgolbeck@umd.edu

ABSTRACT

There are many search and browsing tasks online where relevance scores are not particularly important to the user, but other scores like popularity or average rating can be very informative. If and how these scores are shown varies widely between systems. In this paper, we investigate different methods for visualizing these scores and how they affect user behavior. We conducted a controlled study with 21 subjects who each completed tasks with six different visualization methods. We found that there was no significant difference between the methods with respect to their impact on the user interaction with search results, but that there was a strong preference for having some sort of visualization. We discuss the experiment, results, and design implications that follow from this work.

Author Keywords

information visualization

ACM Classification Keywords

H5.m Information interfaces and presentation: Miscellaneous

INTRODUCTION

In many online systems, users perform a search and the set of items returned are all equally (or similarly) relevant to the search. The user must then browse the results to find what they are looking for. For example, a user may want to find a song and know the name of the artist, but not the song title. A search for the artist in iTunes will return a series of results - all of which may match the query. Similarly, a Netflix user may search for a keyword (e.g. "zombies") and the results will show movies that all match that keyword. Then, users can typically preview the items; e.g. in iTunes, there are 30 second song previews and in Netflix a mouse-over shows a short description and a few statistics about the movie.

The relevance score is not as useful in these tasks as it is in document search, but there are other scores which can be quite helpful. Some systems show these other scores to the user with a visualization, and other systems do not. In iTunes, popularity is shown with a bar (figure 1(a)); user ratings and recommendations, such as those in Netflix or online product review websites, are often shown on a star scale (figure 1(b)); email messages scored by importance are sometimes color coded (figure 1(c)). Yet, not all systems have visualizations. For example, while the iTunes Store shows popularity ratings and allows users to sort by that, the Amazon MP3 site does not show popularity information at all.

Do these visualizations impact user behavior?Furthermore, different domains have their own standards for visualizing ratings or scores associated with items. Do users expect certain types of visualizations in certain domains and does a change from the meme affect the way they interact with the information?

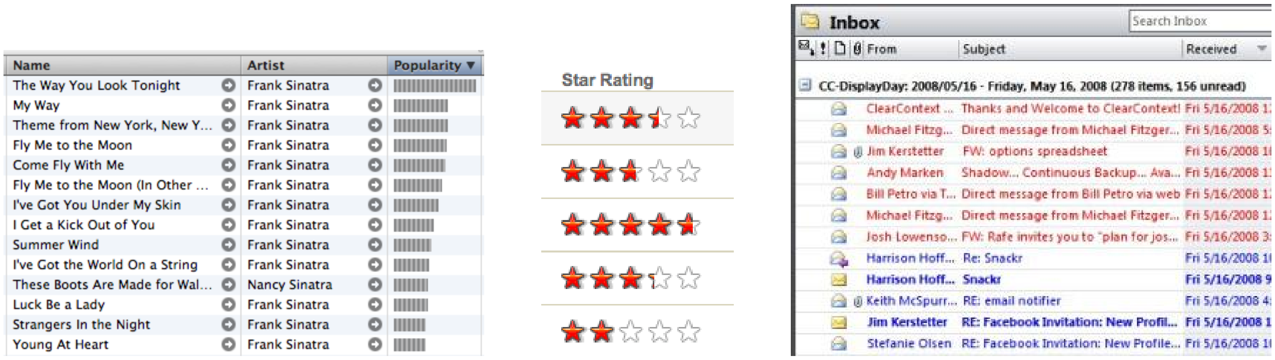
In this paper, we study the effectiveness of four types of visualizations - color coding, bars, colored bars, and stars - and compare them with two controls: no indication of the score, and display of the numerical value. This is tested with a task similar to that described above. Given a set of search results, we ask users how many items they would preview before reformulating or abandoning their search.

Through a user study, we found no significant difference in how many items users would consider, but that there was a strong user preference for having some sort of visualization. In the rest of this paper, we briefly describe related work, then explain our experiment and results, and discuss design implications that follow.

RELATED WORK

There have been many methods for visualizing search results. While we are interested in quantitative descriptors beyond relevance (e.g. popularity, importance, etc.) relevance itself has obviously been the most widely studied quantitative value relating to search results. There are many techniques, but few studies have

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI 2009, April 4 - 9, 2009, Boston, Massachusetts, USA.
Copyright 2009 ACM 978-1-60558-246-7/09/04...$5.00.



Name		Artist		Popularity ▼
The Way You Look Tonight	➲	Frank Sinatra	➲	
My Way	➲	Frank Sinatra	➲	
Theme from New York, New Y...	➲	Frank Sinatra	➲	
Fly Me to the Moon	➲	Frank Sinatra	➲	
Come Fly With Me	➲	Frank Sinatra	➲	
Fly Me to the Moon (In Other ...	➲	Frank Sinatra	➲	
I've Got You Under My Skin	➲	Frank Sinatra	➲	
I Get a Kick Out of You	➲	Frank Sinatra	➲	
Summer Wind	➲	Frank Sinatra	➲	
I've Got the World On a String	➲	Frank Sinatra	➲	
These Boots Are Made for Wal...	➲	Nancy Sinatra	➲	
Luck Be a Lady	➲	Frank Sinatra	➲	
Strangers In the Night	➲	Frank Sinatra	➲	
Young At Heart	➲	Frank Sinatra	➲	

(a)

(b)

Inbox

Search Inbox

From	Subject	Received
CC-DisplayDay: 2008/05/16 - Friday, May 16, 2008 (278 items, 156 unread)		
ClearContext ...	Thanks and Welcome to ClearContext!	Fri 5/16/2008 1:
Michael Fitzg...	Direct message from Michael Fitzger...	Fri 5/16/2008 5:
Jim Kerstetter	FW: options spreadsheet	Fri 5/16/2008 1(
Andy Marken	Shadow... Continuous Backup... Ava...	Fri 5/16/2008 1:
Bill Petro via T...	Direct message from Bill Petro via web	Fri 5/16/2008 1:
Michael Fitzg...	Direct message from Michael Fitzger...	Fri 5/16/2008 1:
Josh Lowenso...	FW: Rafe invites you to "plan for jos...	Fri 5/16/2008 3:
Harrison Hoff...	Re: Snackr	Fri 5/16/2008 1(
Harrison Hoff...	**Snackr**	**Fri 5/16/2008 9**
Keith McSpurr...	RE: email notifier	Fri 5/16/2008 4:
Jim Kerstetter	**RE: Facebook Invitation: New Profil...**	**Fri 5/16/2008 1**
Stefanie Olsen	RE: Facebook Invitation: New Profile...	Fri 5/16/2008 1(

(c)

Figure 1. Some existing visualizations of scored items: (a) Bars indicating song popularity in iTunes, (b) Stars showing recommended ratings for movies in Netflix, (c) Color coding email messages by importance in ClearContext

shown positive results and these techniques have not been widely implemented (see [2] for an excellent review of the literature). Nearly all of the research has focused on using alternative displays to the standard web-view that lists results. Spatial visualization was present in many techniques, but a meta-analysis conducted in 2000 showed no significant benefit from spatial visualization of search results [1].

Bars have frequently been used to visualize quantitative data, generally and with respect to search results specifically. Table Lens is one of the most widely recognized examples of visualizing values as bars, and this proved to be an effective means of seeing patterns in tabular data [4]. A similar use of bars to indicate relevance and other quantitative descriptors of search results was tested with users in [5]. Subjects did not like this view and it was less effective than other views (including a standard web search view). However, this may be due to the layout where the bars were shown together at the top of the page with no text preview and thus out of context with the content of the search results.

In this work, we are not interested specifically in what type of visualization makes users more effective. Instead, we are interested in what impact different visualizations have on users' perception of relevance and if this changes based on context. We did not find any previous work specifically comparing different types of simple visualization (bars, stars, color coding).

EXPERIMENT

We conducted a controlled user study to answer one major research question: What impact do visualization methods have on how many items a user will consider before stopping the search or reformulating the query? We also looked at the impact of these visualization methods on the data divided by the structure of the underlying data and by the domain to which the task applied.

Visualizations

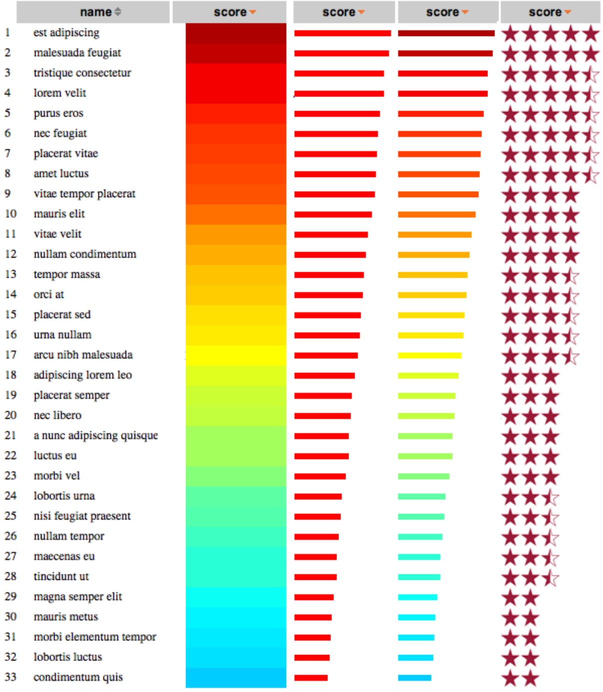


Figure 2. Visualizing scored items: (a) Bars indicating song popularity in iTunes, (b) Stars showing recommended ratings for movies in Netflix, (c) Color coding email messages by importance in ClearContext

Four visualizations and two controls were used in this experiment:

Visualizations

- Color coding - a color bar indicates the score of the item. Many color schemes could be used but ours ranged across the spectrum from blue (low scores) to red (high scores)
- Bars - the width of the bar indicates the magnitude of the score
- Color coded bars - this feature combines the color coding scheme with the bars to provide two indica-

tions of the score
- Stars - a frequently used meme for ratings systems, ratings are mapped to a star model with 0 to 5 stars including half star ratings

Each visualization method is illustrated in figure 2.

In addition, we used two controls. In one, no indication of the score was shown. In the other, the numerical value of the score itself was shown as a percentage next to each item. In all cases, the items were shown to the user sorted from highest scored to lowest scored.

Setup

Subjects were asked to complete one task in a variety of conditions: indicate how many items they would examine before giving up or reformulating their search. This was presented in two domains: searching for files generally (which may include images, MP3s, documents, people in a social network, etc.) and searching for rated items. The instructions were presented as follows:

- Imagine you have searched for a file and the system has returned this list for you, sorted by the popularity/importance from highest to lowest. How many files would you preview (reading snippets or opening) without finding a match before you would stop looking or try a new search?
- Imagine you have searched for recommended movies and the system has returned this list for you, sorted by the average rating from highest to lowest. How many movies would you preview (reading descriptions or watching trailers) without finding a match before you would stop looking or try a new search?

Subjects indicated the last item they would examine by dragging a slider to highlight that item.

For each of these two domains, subjects were shown a series of lists of scored items, each displayed with one of the visualization methods or with a control method. The item names were intentionally meaningless to prevent any domain knowledge from becoming a factor.

For each visualization method, two ratings distributions were used. One used an approximately linear decrease in scores with slight random noise added to make the values look realistic. This linear distribution is shown in the visualizations in figure 2. The second distribution had a sharp drop off with the first three items scored over 0.85 and the rest below 0.6 descending on a linear scale.

This created a total of 24 conditions: two domains × six visualization methods × two score distributions. Each subject saw all 24 conditions.

The subjects were then asked post-test questions:

1. Please score how much you liked each of the methods for visualizing movies where 1 means you did not like it at all and 5 means you liked it very much. (Each visualization method is shown with a 5 point likert scale)
2. Please score how useful you found each of the methods for visualizing files where 1 means it was not at all useful and and 5 means it was very useful. (Each visualization method is shown with a 5 point likert scale)
3. Do you think one of the visualization methods is best to use in every application (i.e. it's always better than the alternatives)?
   (a) If yes, which one?
   (b) If no, do you think one visualization method is good for some applications while another method is better for a different application? If yes, please explain.

Subjects

We deployed the experiment online. Twenty-one (10 female, 11 male) subjects finished all the conditions. Most female subjects reported weekly use of rating web sites. Most male subjects reported daily use of such web sites. Most users are in the 25-35 age group.

Results

We used two underlying datasets. One had a sharp drop in the scores between the 3rd and 4th item and the other had a linear decrease in the scores. Over all the visualization methods, we found that users chose significantly different cutoff points in these data sets. With the sharp drop off, the average cutoff point was 9.73 vs 13.75 for the linear data. A two-tailed t-test showed this difference was significant for $p < 0.001$.

Overall, we found no significant difference in the selected cutoff point based on the visualization methods. This analysis included the control with no visualization or indication of the score. Breaking the data down, the negative results persisted. Within each task (searching for files vs searching for movies) we found no significant difference between the visualization methods. Similarly, within each dataset (linear vs sharp drop off) there was no significant difference between visualization methods.

The subjects' responses provided some interesting insights. Sixteen of our subjects responded to the post-test questions. We found significant differences in how much they liked each visualization type and how useful they found it to be. An ANOVA indicated significant differences among the subjects' ratings of how much they liked each visualization method (F(5,90) = 10.05, $p < 0.001$) and how useful they found each visualization method (F(5,90)=6.37, $p < 0.001$). As shown in figure , subjects rated the control condition of no visualization method lowest in terms of how much it was liked and how useful it was. The figure shows ratings for both questions with a 95% confidence interval indicated by the error bars.

T-tests show that the no-visualization control is rated

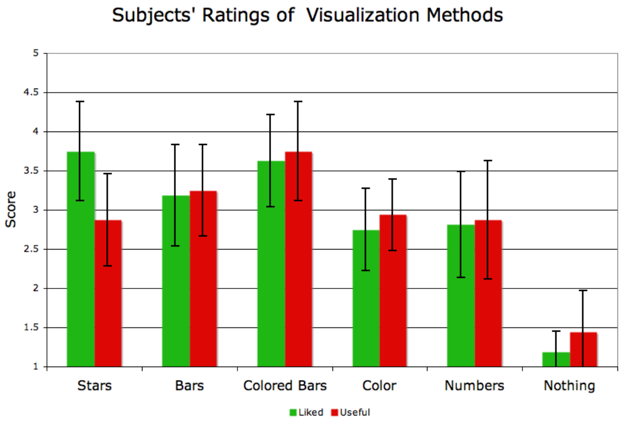


Figure 3. Subjects' preferences for the different visualization methods. Green bars indicate how much they liked the method and red bars indicate how useful they found the method to be. Error bars indicate 95% confidence intervals.

Table 1. Subjects' self-reported preferred visualization methods

No	Colored				
Preference	Bars	Stars	Numbers	Color	Bars
6	4	2	2	1	1

significantly lower than all other visualization methods ($p < 0.001$) on both the "like" and "useful" questions.

In the final question where users were asked if they thought there was one visualization method that was always better than others. Results were mixed, as shown in table 1. Among subjects who had no preference, four answered the follow-up question and all four indicated that they felt stars were the most appropriate visualization for movie ratings.

DISCUSSION

Most subjects interviewed reported that they had an "internal" cutoff point which they would use when the scores seem unreliable, or no obvious cutoff point in the distribution can be found. Since we found no difference between the visualization methods for our tasks, it appears that this pre-determined, intuitive cutoff point is more important than if or how users see scores.

However, despite the fact that users did not make their choices any differently with or without visualizations, they reported that they liked all the visualizations better than no visualization at all, and that they found all of the visualizations to be more useful than no visualization at all. Stars and Colored Bars were the most liked visualization methods, but stars' usefulness rating was much lower. Through these ratings, subjects indicate that they may like a visualization that they know is not particularly useful to them. This echoes results from other information visualization studies where user preferences for a visualization do not necessarily correspond to an improvement in the interaction [3].

These results lead to design guidelines. While previous work has not shown an increase in performance when scores of search results are visualized, and our results showed no significant differences in behavior based on visualization, there is a clear user preference for having some indication of the score. Some existing systems show scores (e.g. the popularity bars in the iTunes store) while other similar systems do not (e.g. the MP3 store at Amazon.com). Since subjects preferred some method of visualization to nothing, using visualization is likely to improve user satisfaction with these systems, even while it does not change the users' behavior.

CONCLUSION AND FUTURE WORK

Many systems have scores, like popularity, average rating, or importance, that may be valuable to users. Some systems display this information and others do not. In this study, we used several visualization methods a investigated how they impacted the way users interact with search results. We found no significant difference in user behavior between any of the visualization methods or without visualization at all. However, users indicated a strong preference for having some sort of visualization over none. They rated that they liked the visualizations and that the visualizations were more useful than having no data visible.

This suggests that users may be more likely to use systems that show these scores than those that do not, even though their interactions might be similar. When there is competition between systems (e.g. iTunes vs. Amazon MP3), user preference for the interface may be an important factor in what customers use. Further research will be necessary to understand if there are any other benefits or drawbacks to these visualizations that should be considered.

REFERENCES

1. Chen, C., and Yu, Y. Empirical studies of information visualization: a meta-analysis. *International Journal of Human Computer Studies 53*, 5 (2000), 851.

2. Hearst, M. A. *Search User Interfaces.* Cambridge University Press, 2009.

3. Hornbaek, K., Bederson, B. B., and Plaisant, C. Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM Trans. Comput.-Hum. Interact. 9*, 4 (2002), 362–389.

4. Rao, R., and Card, S. K. The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 1994), ACM, pp. 318–322.

5. Reiterer, H., Tullius, G., and Mann, T. Insyder: a content-based visual-information-seeking system for the web. *International Journal on Digital Libraries 5*, 1 (2005), 25–41.