

WISDNA: An Information Visualization Paradigm for XML

Ricardo Baeza-Yates, Ricardo Lemus

Center for Web Research, CS Dept.

Universidad de Chile, Santiago, Chile

Dulce Ponceleon, Savitha Srinivasan

IBM Almaden Research Labs.

650 Harry Rd, California, USA

INTRODUCTION

With ever-increasing amounts of data in repositories, finding the information a user needs becomes increasingly challenging. When users approach an information access system they often do not have a clear understanding of how to access the precise information they need. Additionally, it is not rare that users have only a vague notion of what they are looking for, hence the need for an exploratory navigation tool beyond a static list of search results. A model used in practice is an iterative search-and-browse cycle to retrieve relevant information. The role of the user interface in understanding search results and providing relevant metadata in a compact manner is a key component of this iterative process. In this context we present WISDNA as an information visualization paradigm for XML data from web information services.

According to the Department of Commerce U.S. online retail sales topped \$32 billion in 2002, a 19 percent increase over the previous year. In 2002, consumers continue to embrace online commerce, despite the economic slowdown as exemplified by Amazon's sales of books, music, and videos rose by 8 percent over the previous year. A key aspect of online storefronts is partnerships with brick-and-mortar brands and expertise to operate joint online stores. Having perfected ways to put vast inventories at customers' fingertips, online stores continue to rely on familiar models for visualization based on a linear list of search results. However, with increasing complexity in the categorization and attributes of the merchandise they sell, users need more effective tools to navigate this information. Unfortunately, most web-sites provide the traditional database approach of submitting a query, and packaging the response in an HTML page where typically further exploration of search results is not an option [Shafer00]. Web applications such as e-commerce applications need a more flexible metaphor.

We propose a visualization and navigational tool based on Context Lenses and MovieDNA [Russell00, Ponceleon01] to address the need to support exploratory navigation that seamlessly integrates querying with browsing. We present linear and hierarchical versions of WISDNA. It is a dynamic visualization that provides concise representation of search results (XML data) for web information services. It is particularly suited

to provide feedback on associated metadata and relationships with previous user searches. We illustrate our work in the context of online retail applications.

WISDNA

The WISDNA (Web Information Services DNA) technique is a visualization technique for XML data. It builds upon the visualization techniques of *brushing and linking*, *focus-plus-context*, *magic lenses* and the use of *animation* to retain context [Hearst99]. The visualization adheres to basic design principles of providing users with feedback about the relationship between their query specification and documents retrieved, about relationships among retrieved documents, and about relationships between retrieved documents and metadata describing collections. Like in the MovieDNA, we have borrowed the DNA term to just indicate the visual similarity to DNA prints. There is no connection on the user's ability to read DNA prints and the ability to interpret WISDNA visualizations.

Consider the task of searching the Amazon online retail store for content that is available in the following media types: books, classical music, DVD, popular music, video and computer games. A user has two options. One option is to search the entire set of media types (including magazines, software, images etc.) and to sift through the search results for the media types of interest. A second option is to search each media type one at a time for topics of interest. In both cases, a linear list of search results must be browsed, remembered and compared with the rest of the result set. An even more challenging scenario is to compare the results obtained from different search sessions, without having to re-create previous searches. It is necessary to provide a solution beyond adding interesting items to the wish list or bookmarking relevant items.

Brushing is an interaction style that provides at-glance detailed information at low cost. Here cost reflects user time and number of user actions [Russell93]. WISDNA is designed to show relationships between the specific and the whole at once while exploring the search results. Seeing an abstraction of *all* the results provides global context while seeing a specific image or relevant metadata or feature associated to a specific item provides a focus view. Therefore WISDNA is a

focus+context visualization technique. It provides the added benefit of reducing working memory load by providing mechanisms for keeping track of choices made during the search process, allowing users to return to temporarily abandoned strategies, jump from one strategy to the next, and retain information and context across search sessions.

At glance, the WISDNA visualization is very similar to a vertical TileBar [Hearst00] however the introduction of brushing makes WISDNA non-static. The WISDNA context lens is a grid containing as many rows as there are search results. The sorting of the rows can be done in different ways. We have chosen to the ranking order

provided by Amazon, however standard IR ranking might not be the best in this context, instead clustering of results by certain attributes (such as categories) might render an more readable WISDNA. The columns represent features and/or attributes of interest. Each entry in the WISDNA indicates the presence/absence of an attribute of interested (column) in a specific item (row). As an example we use the Amazon Web Services specification, so the relevant attributes are the different product categories (i.e., books, DVD, video, etc), media types (i.e., paperback, hard cover, audio-cassette, audio CD, VHS-Tape, etc.) and other attributes of interest such as availability, customer ranking, average price, etc. Different colors or a gray scale can indicate relevance, scoring, and multiple occurrences of attributes in a search result. For example, color could be use to immediately draw attention to those items with price below certain threshold.

When the user moves the mouse over the search results (i.e. brushing) every line in the DNA triggers the display of a fold-out window to the right of the corresponding line. This window displays detailed metadata on the search result in question. Figure 1 shows the search results displayed for the query “Harry Potter” using a linear DNA, and the metadata that folds out as a result of brushing over item 7. By linear DNA we mean there is only one grid, and the metadata folds out when brushing a row.

With this instance of WISDNA, that we call AmazonDNA, users can easily gain an overview of the search results in the context of the media types of interest. An important feature is that the fold out window does not cover the DNA grid. This allows the user to study the in-focus information while maintaining a sense of context

and location in the search results. Analogously to the work in MovieDNA, we can define a hierarchical WISDNA where there is more than one level of information. Each level is represented as a DNA grid. For example, a linear DNA can shrink or aggregated Figure 2 shows an example of an aggregated DNA for a book query. In this case the first level DNA is trivially one row, indicating that there were hits only in the book category with only two attributes of interest (paperback and hard cover). Brushing over a row of this first level triggers the second level DNA, where more detailed information is displayed such as the ISBN number and price (although not shown in this picture). Finally when brushing over this 2nd level DNA the metadata associated folds out (in this case the image of the book). Finally, Figure 3 shows a WISDNA example for all product lines. The first level shows attributes common to all products, the second level shows attributes of a specific product line, and finally the image shows the metadata associated to the fifth result (a software game).

WISDNA is different from other visualizations because it combines global viewing, detailed viewing and filtering in one interaction. The grids provide global view, the choice of attributes/columns enables filtering and during brushing only *one* item is revealed in full detail without interfering with the global view.

Search for **Baeza Yates** in Amazon.com WebService

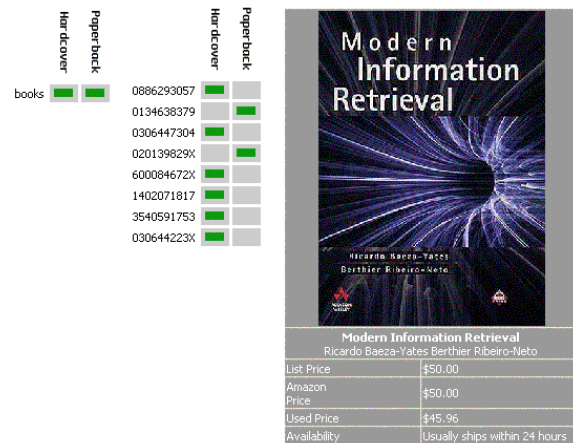


Figure 2: Hierarchical WISDNA, Two-Level Search Results for “Baeza-Yates” in books category.

SYSTEM IMPLEMENTATION

Our first prototype was done in Perl using the SOAP::Lite module, executing under mod_perl (but can also use cgi-bin). The user interface, available in <http://soap.lemus.cl/>, allows typing the

query, the product category, and the number of results per page wanted in the DNA. We use SOAP to query the Amazon Web service according to their specifications, and processing the answer to obtain the desired data. If all the products are chosen, the system queries the six available Amazon categories. The display interface uses plain HTML to show the first DNA level and layered HTML for the other two levels. A second prototype uses Java to improve the DNA visualization, showing all the concepts behind our idea and is very light. Most of the answer delay is due to the Amazon web service delay.

CONCLUSION AND FUTURE WORK

We have presented a tool to visualize information available in many Web services which opens the door to better user interfaces, in particular for e-commerce sites with large databases. Results of user studies performed on previous Content Lenses work show that brushing provides a qualitatively interaction style that is more powerful than a series of selections in a hierarchical interface [Russell00]. The key difference is the delayed commitment to a particular item until they are certain about their interest. For the AmazonDNA other instances of relevant metadata can be defined. For example, in order to incorporate the information about "*people that bought this title also bought the following title*", one can build a storyboard containing thumbnails of such titles and include it as the fold-out metadata. The tool presented here can be complemented with similar DNA information of

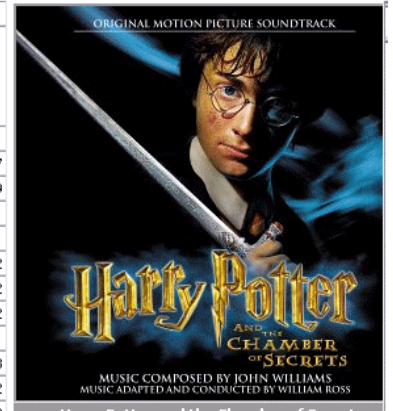
the user history, such as products in the shopping cart, queries and links used, etc. Our current prototype handles two levels, but the generic idea is to navigate in arbitrary XML data in the future. This is the main difference between visualizing relational data and XML data.

REFERENCES

- [Dieberger00] Dieberger, A. and Russell, D.M., Context Lenses – Document Visualization and Navigation Tools for Rapid Access to Detail, in Proceedings of ACM CHI 2000.
- [Hearst99] Hearst, M.A., User Interfaces and Visualization, edited by Baeza-Yates, R., and Ribeiro-Neto, B., Modern Information Retrieval, Addison-Wesley, ACM Press, New York, January 1999.
- [PonceLeon01] Ponceleon D., and Dieberger A., Hierarchical Brushing in a Collection of Video Data, in HICSS'34 (Hawaii International Conference on Systems Science), MiniTrack on Video in the office, Maui, HI, January 2001.
- [Rao94] Rao, R., and Card, S. K., The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information. CHI'94, Boston, ACM Press: 318-322.
- [Russell94] Russell, D. M., et al., The Cost Structure of Sensemaking , in InterCHI'03, 1993, Amsterdam, p. 269-276.
- [Shaer00] Shafer J., and Agrawal R., Continuous Querying in Database Centric Web Applications. 9th Int'l World Wide Web Conf., Amsterdam, May 2000.

Search for **harry potter** in Amazon.com WebService

#	Product	Product Line							Media Type										Average Customer Rank	Sales Rank
		Books	Classical Music	DVD	Popular Music	Video	Computer & Video Games	Paperback	Hardcover	Audio Cassette	Unknown Binding	Audio CD	DVD	VHS Tape	Video Game	CD-ROM	Used Price	Collectible Price		
1	043935806X	■																■	1	
2	0439249546	■														■	■	■	893	4.67
3	0972393617	■						■										■	80	4.79
4	B000004433		■								■							■	81,541	5
5	B0000005L2		■								■							■	127,351	5
6	B00008DDXC			■								■						■	1	4.42
7	B00008DDXL			■								■						■	10	4.42
8	B00003CX11			■								■						■	114	4.12
9	B000083MF8				■						■							■	316,706	
10	B00006IR55				■						■							■	819	4.73
11	B00005OWIU				■						■							■	1,788	4.52
12	B00008DDXB					■							■					■	7	4.42
13	B00003CX10					■							■					■	206	4.12
14	0780614097					■							■					■	5,080	4.45
15	B00006F7RS						■							■				■	79	4.3
16	B00006JB55							■						■				■	167	3.7
17	B000061JJM							■							■			■	588	3.88



ORIGINAL MOTION PICTURE SOUNDTRACK

Harry Potter
AND THE CHAMBER OF SECRETS

MUSIC COMPOSED BY JOHN WILLIAMS
MUSIC ADAPTED AND CONDUCTED BY WILLIAM ROSS

Harry Potter and the Chamber of Secrets
J. K. Rowling Mary GrandPré

List Price	\$18.98
Amazon Price	\$13.99
Used Price	\$8.38
Availability	Usually ships within 24 hours

Figure 1: Linear WISDNA, Search results for “Harry Potter”.

Search for **Depeche mode** in Amazon.com WebService

	VHS Tape	Video Game	CD-ROM	DVD	Audio Cassette	Paperback	Audio CD	VHS Tape	Video Game	CD-ROM	DVD	Audio Cassette	Paperback	Audio CD
vhs	■													
videogames		■	■						■					
dvd				■						■				
books					■	■				■				
music							■			■				
classical							■			■				



GAME BOY ADVANCE

Moto Racer
advance

LICENSED BY
Nintendo **Ubi Soft**

Moto Racer

List Price	
Amazon Price	\$29.99
Used Price	
Availability	This item is not stocked or has been discontinued.

Figure 4: AmazonDNA , Hierarchical WISDNA, Two-Level Search Results for “Depeche Moto”.