

# Visualizing Reputation and Recommendation in Scientific Literature

**Pearl Pu**

Human Computer Interaction Group  
Global Computing Center  
Swiss Institute of Technology Lausanne  
CH-1015 Ecublens, Switzerland  
Pearl.pu@epfl.ch

**Punit Gupta**

Indian Institute of Technology Guwahati  
Guwahati-781039,  
Assam India.  
punitkg@yahoo.com

## ABSTRACT

In this paper, we investigate the architecture for visual social navigation systems, especially those used in scientific literature. We identify the requirements for effective visualization of reputation and recommendation, two notions that underlie most social navigation systems. We present several designs of visual emphasis techniques addressing the dimension and scales of reputation, and two sets of icons for effectively distinguishing recommended items. The results of two user studies are then discussed, which provided validation for our non-linear emphasis methods to represent reputation, and for choosing optimal icons for recommended items. Based on these findings, we present a re-design for the NEC Research Institute ResearchIndex (Citeseer), a concrete case of reputation based recommendation system.

## KEYWORDS

Visual reputation and recommendation systems, information visualization, social navigation, emphasis techniques, scientific literature search.

## INTRODUCTION

From maps etched on cave walls indicating shared knowledge of hunting trips in ancient times, to dog-eared books in libraries marking their popularity, or smoothed trails in forests indicating paths leading somewhere, social environments have enabled humans to accomplish complex tasks based on information gathered collaboratively. Such information is often presented in the forms of objects bearing signs of how they have been used by others, whether as maps, tear-and-wear signs on books, or footprints [10]. At the same time, such information tends to be very visual. Partly it is because of the *navigational* nature of user tasks in such social environments (hunters must

quickly go on hunting, library users must quickly decide which books to read), thus requiring a quick-to-perceive medium for communication and understanding. Partly it is due to the fact that traditional social spaces are physical and therefore intrinsically visual. Today, however, we go to the World Wide Web to find out where to dine, which books or CDs to buy, or where to go for vacations [1]. Digital spaces are replacing traditional social environments, and employ mechanisms of collaborative filtering and recommendation systems to enable users to navigate and find useful information in these spaces [12]. Unfortunately, digital spaces do not age, nor are they physical in nature. Most information concerning reputation, and recommendation are text based, forcing users to “read” rather than affording them to “see” information, thus making it hard for them to quickly perceive evidence, spot and retain trends, and patterns. “Read” and “see” differ in terms of the two ways in which the human visual system processes information, one being controlled and other automatic processing [4][8].

We are concerned with investigating an architecture for visual social navigation systems, where “social translucency” is mapped to visual representation and therefore augments users capabilities in their navigation skills.

We aim at developing methods to effectively communicate reputation and recommendation information to achieve the following objectives:

1. provide a social space that resembles the multi-dimensional physical space in which we habit;
2. provide distinguishable and intuitive visual signs for reputation, enabling searchers to quickly decide which items to pursue (decision making in navigation), without making decisions for them;
3. provide visual representation for recommended items to enable users to “see” items in the right categories, rather than “read” them;
4. provide users with quick learning curves to understand and retain the meanings of signs in such environments.

We have chosen the scientific literature as our case study because it is rich in information structure, there exists well established computation algorithms to calculate reputation and recommendation information, and finally digital libraries based on social navigation is gaining popularity and calls for more visual interfaces to increase user experience and augment task efficiency.

### SURVEY OF EXISTING VISUAL CUES FOR REPUTATION

We distinguish our work from the computation of reputation and recommendation, and the use of reputation in multi-agent systems, although the results of our work can be applied to any system where reputation and recommendation values may be visualized for end users.

We adopt the definition of reputation as being the accumulated scale of opinions of products or persons from a significant population of people. Social reputation refers to such treatment of reputation within a defined social group or network for which the user belongs to [7]. In general, a reputation is a quantitative score such as 4 stars, or an ordinal scale such as very good, good, or poor. Some reputation notions are context dependent in that a product such as a car can be rated in terms of its seat capacity, gasoline efficiency, engine power, customer service support, and etc. We also use the term reputation, ranking, and rating interchangeably.

Most visualization of reputation used in current social navigation systems is quite limited. According to our survey, the following schemes are used:

**Ordered List** Link analysis and page ranking algorithms employed by some of the search engines can be considered as computing the reputation of web sites by analyzing the site's link popularity [11][3]. Results are presented in a list arranged in order of the ranks, with most popular ones first. A result page contains around 10 website links.

Traditional methods used in guidebooks, and more recently in e-commerce sites (Amazon and eBay), or intermediaries such as e-opinion use stars, dollar signs, text labels, or other symbols for rating and ranking.

**Stars** (restaurant guides, electronic devices): Rating is represented by a number of stars linearly arranged. Some sites also gray out stars to enhance the comparative viewing of low and high scores such as used in Figure 1.



Figure 1. Stars are used to present the rating of songs.

**Dollar signs:** Rating is represented by a number of dollar signs indicating the price of the item.

**Textual labeling** (new, popular, hot, bestsellers): Textual labeling can be used alone, or is sometimes combined with symbols to provide additional reputation (Figure 2).

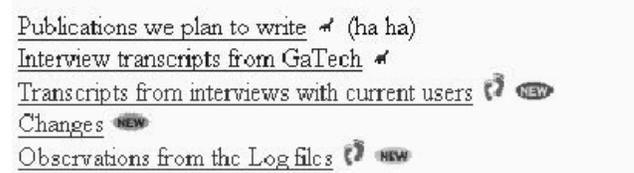


Figure 2. Textual labels show reputation, where red colored label in the middle row is the hottest item among them. Items are arranged in order of ratings with highest rated item on the top.

**Multi-criteria rating:** Systems such as Epinions ([www.epinions.com](http://www.epinions.com)) use featured ratings, where multiple ratings represent the different aspects of a product (Figure 3). This is particularly useful for complex products where the judgment of an entity depends on its various features.

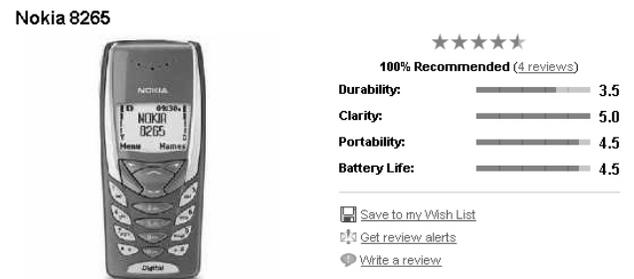


Figure 3. Bars are used to show ratings of different features such as durability, clarity, portability etc.

### RECOMMENDATION

Recommendation, on the other hand, is computed information to help users to find similar items and related items based on knowledge of users' profiles such as their interests, criteria, and constraints. There are two basic forms of recommendations. An explicit recommendation is often provided by a search engine which returns for example a list of restaurant names when a query has been specified. An implicit recommendation is, on the other hand, given to users when related items are thought to be "interesting" for them. An example is the "customers who bought this book also bought" recommendation based on collaborative filtering schemes used in Amazon. Visualization work in this area has been very limited, mostly employing textual labeling.

Researchers of a number of areas have been working on the calculation and use of reputation and recommendation. For a survey, please refer to [14][9]. We focus our investigation of reputation and recommendation in scientific literature, sometimes called the scientometrics.

Reputation and recommendation are different concepts. The former stresses on the overall calculation of statistically accumulated score, while the later takes into account

also the personalization aspect of information. In a social navigational environment, it is important to connect the two concepts to enable users to “see” reputation in the recommended items, justifying a navigator’s pursuit of information. Finally it is important to study users reaction to the effectiveness of visual interface designs for such environments.

## REPUTATION IN SCIENTIFIC LITERATURE

### Dimensionality of reputation

Scientific literature thus falls into the category of items for which featured reputation mechanisms must be used. This calls for the simultaneous representation of multiple criteria, and visualization in high dimensions of data relationships. One method (such as used in Figure 3) is to order the literature items in a list, with parallel bar charts indicating the ratings of individual features, for example, citation number, author reputation, lieu of publication, year, the number of downloads, and etc. However, the total number of items that can fit into a page will be compromised.

For the particular digital library (Citeseer) where we are investigating our design issues, the reputation information is limited to the number of citations and the year of publication. Data on the popularity of an author can be obtained, but rather separately from the cited articles. Being limiting to the capabilities of browsers and the fact we want to display as many titles as possible for an optimal context for comparison, we have therefore decided to employ a two-dimensional visualization to principally representation the number of citations and the year of publication.

### Scales of reputation

The range of values used to represent the scale of citation can vary in a wide range in scientific literature. The number goes from 0 (only self citation) to as high as 10,000 (record at Citeseer is 10473 for authors, 1198 for Documents).

To understand the pattern and ranges of citations, we examined a set of 100 documents at Citeseer, and plotted a graph on the frequency of different scales that appear in the collection. It turns out a majority (60%) of papers are cited only up to 3 times, 25% are cited between 4 and 30, 10% are cited between 31-100 and 5% are cited more than 100. This graph (where the x axis represents the number of citations and the y axis the number of documents) starts with many documents cluttered around 0-30 and trails off to few documents with high citation numbers. The high density in the low ranges of citation and the wide gap between low and high number of citations thus calls for some non-linear mapping techniques between the actual citation value and what can be visually displayed. In particular, we must distinguish the scales conforming to our study of statistical frequencies of citation numbers. That is, a wide range of values must be

modeled, while at the same time the mapping will result in a range of 10 distinct visual ranges. We have opted for the slow-in slow-out method [13], which magnifies the highly dense areas, and de-magnifies in the low distribution area (Figure 4).

### Equation 1

$$Y = \frac{(d+1)*X}{d*X+1}$$

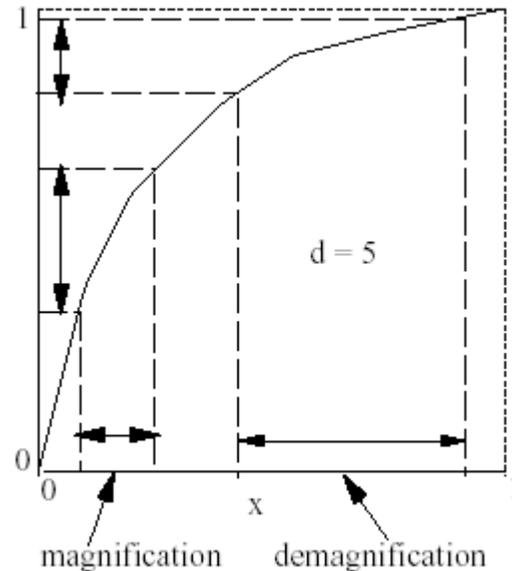


Figure 4. The graph for equation 1 showing regions of magnification and demagnification

### Initial designs for citations

The first design issue is to determine the most effective mapping between citation scales and visual values using the slow-in and slow-out function. Let Max represent the highest number of citation of a given result page. We consider the following choices:

1. Font size– the values between 0 and Max are mapped to a range between 0 to 1, using 5 distinct font sizes (Figure 5 where Max = 185);
2. Stars – the values between 0 and Max are mapped to a range between 0 and 1, using a set of 10 stars linearly arranged (Figure 6 where Max=185);
3. Font size + weight + shading – the values between 0 and Max are mapped to a range between 0 and 1, using 7 scales of combined effects (Figure 7 where Max=185). Here bold-faced characters are heavier than non-bold faced one. The shading of characters refers to the gray scale of their fonts.

### Non-linear visual values

Stars, or a series of any symbols linearly arranged, are linear visual signs to render quantitative values. According to [2,5,6,8], it is easier for humans to judge the comparative values represented by lines, than by other forms

such as area or volume. However, since our underlying scales vary non-linearly, we have decided to explore visual forms based on area and weights so that items of distinguishing qualities will stand out.

Title	Author	Year
<b>Oxford University Press (context)</b>	<b>Dawkins</b>	<b>1976</b>
<b>Information retrieval</b>	<b>vanRijsbergen</b>	<b>1979</b>
<i>Journal of Experimental Psychology: Learning (context)</i>	Anderson, Pirolli	1984
<i>Princeton University Press (context)</i>	Stephens, Krebs	1986
<i>Phase transitions in artificial intelligence systems (context)</i>	Huberman, Hogg	1987
<i>Human memory: An adaptive perspective (context)</i>	Anderson, Milson	1989
<b>The adaptive character of thought (context)</b>	<b>Anderson</b>	<b>1990</b>
<i>Identifying aggregates in hypertext structures (context)</i>	Botafogo, Schneideman	1991
<i>An Object-Oriented Architecture for Text Retrieval</i>	Cutting, Pedersen et al.	1991
<i>Evolutionary ecology and human behavior (context)</i>	Smith	1992
<b>Information visualization using 3D interactive animation (context)</b>	<b>Robertson, Card et al</b>	<b>1993</b>
<b>Rules of the mind (context)</b>	<b>Anderson</b>	<b>1993</b>
<i>Indecor the internet (context)</i>	Turber	1995
<i>Results from the Third WWW Survey (context)</i>	Pitkow, Johow	1995
<i>Information foraging in information access environments (context)</i>	Pirolli, Card	1995
<b>Visualizing complex hypermedia networks through multiple hie..</b>	<b>Mukherjea, Foley et al</b>	<b>1995</b>
<b>Characterizing browsing strategies in the World-Wide Web</b>	<b>Catledge, Pitkow</b>	<b>1995</b>
<i>Gather browsing communicates the topic structure of a very</i>	Pirolli, Schank et al.	1996
<i>The WebBook and the Web Forager: An Information Workspace</i>	Card, Robertson et al.	1996

Figure 5. Documents have been visually rendered by variations on font size. Documents in bigger fonts correspond to higher numbers of citations.

Title	Author	Year
<b>Oxford University Press (context)</b> *****	Dawkins	1976
<b>Information retrieval</b> *****	vanRijsbergen	1979
<i>Journal of Experimental Psychology: Learning (context)</i> *	Anderson, Pirolli	1984
<i>Princeton University Press (context)</i> *	Stephens, Krebs	1986
<i>Phase transitions in artificial intelligence systems (context)</i> ****	Huberman, Hogg	1987
<i>Human memory: An adaptive perspective (context)</i> *	Anderson, Milson	1989
<b>The adaptive character of thought (context)</b> *****	Anderson	1990
<i>Identifying aggregates in hypertext structures (context)</i> *	Botafogo, Schneideman	1991
<i>An Object-Oriented Architecture for Text Retrieval</i> ***	Cutting, Pedersen et al.	1991
<i>Evolutionary ecology and human behavior (context)</i> *	Smith	1992

Figure 6. A variation using stars is used to show citation numbers.

Title	Author	Year
<b>Oxford University Press</b>	<b>Dawkins</b>	<b>1976</b>
<b>Information retrieval</b>	<b>vanRijsbergen</b>	<b>1979</b>
<i>Journal of Experimental Psychology: Learning (context)</i>	Anderson, Pirolli	1984
<i>Princeton University Press (context)</i>	Stephens, Krebs	1986
<i>Phase transitions in artificial intelligence systems (context)</i>	Huberman, Hogg	1987
<i>Human memory: An adaptive perspective (context)</i>	Anderson, Milson	1989
<b>The adaptive character of thought (context)</b>	<b>Anderson</b>	<b>1990</b>
<i>Identifying aggregates in hypertext structures (context)</i>	Botafogo, Schneideman	1991
<i>An Object-Oriented Architecture for Text Retrieval</i>	Cutting, Pedersen et al.	1991
<i>Evolutionary ecology and human behavior (context)</i>	Smith	1992
<b>Information visualization using 3D interactive animation (context)</b>	<b>Robertson, Card et al</b>	<b>1993</b>
<b>Rules of the mind (context)</b>	<b>Anderson</b>	<b>1993</b>
<i>Indecor the internet (context)</i>	Turber	1995
<i>Results from the Third WWW Survey (context)</i>	Pitkow, Johow	1995
<i>Information foraging in information access environments (context)</i>	Pirolli, Card	1995
<b>Visualizing complex hypermedia networks through multiple hie..</b>	<b>Mukherjea, Foley et al</b>	<b>1995</b>
<b>Characterizing browsing strategies in the World-Wide Web</b>	<b>Catledge, Pitkow</b>	<b>1995</b>
<i>Gather browsing communicates the topic structure of a very</i>	Pirolli, Schank et al.	1996
<i>The WebBook and the Web Forager: An Information Workspace</i>	Card, Robertson et al.	1996

Figure 7. The variation using font size + shading + weight

### Showing details

Before our detailed user studies, we conducted an interview about our various designs. One comment consistently made by our users is that in addition to the visual cues, it is informative to attach the detailed citation number next to each title. In our user study, we thus have included 5 sets of interface designs, two of them emphasizing the additional use of detail.

Table 1: Mapping from a range of citation numbers to each of the three variations of visual implementations: font size, stars, font size + weight + shading.

Citations	Font size	Stars	Font size +Weight +Shading.
(136-151]	Ex	*****	Ex
(121-136]	Ex	*****	Ex
(105-121]		*****	Ex
(91 - 105]	Ex	*****	Ex
(76 - 91]		*****	
(60 - 76]		*****	
(45 - 60]	Ex	****	Ex
(30 - 45]		***	Ex
(15 - 30]		**	
[0 - 15]	Ex	*	Ex

### User studies

The goal of this empirical study was to survey the 5 designs and determine the most effective non-linear mapping of reputation scales. A two-dimensional display is employed where the documents are order by the their year of publication, and each document is visually rendered to show its citation number.

### Setup of experiment:

The five sets of designs given to our subjects are: 1- Font sizes, 2 - Stars, 3- Stars + details, 4- Font sizes + shading (weight), 5- Font sizes + shading (weight) + details

### Hypothesis:

1. The most effective method to represent the rankings of search results is using font sizes, weight and shading of the fonts.
2. Adding details (actual number of citations) enhances effectiveness.

### Subject groups:

Subjects were divided into two groups: the first group is comprised of undergraduate students (very little research experience) and the second of graduate students (research

oriented). They belong to different demographic backgrounds (from India, China, Switzerland, USA, Sweden, Tunisia) and have different academic specialization such as computer science, communication systems, micro engineering, civil engineering and mechanical engineering.

### Questionnaires

Study participants were provided with the following hypothetical situation:

“Imagine that an article (article of interest) has been submitted to a database of scientific literatures. Two sets of results are displayed: those articles citing the article of interest (on the left column), and those articles being cited by the article of interest (on the right). Articles in both columns are sorted according to the year of publication, from the most recent closest to article of interest, which is located in the middle, to least recent ones on the top. The reputation of each article is the total number of citations for that article.”

**User tasks:** each user receives 5 sets of interface designs where reputation information has been rendered using a combination and variation of font size, stars, and font size + shade; users are asked to pretend to “select 5 articles in relation to the current article to further their reading;” they are asked to mark yes to the interface designs (marking more than once is allowed), if it allows effective scanning of the search results in order to make their selections.

**Results:** Results are quite different for the two subject groups, which were shown in Table 2 and 3 below.

(Fs: Font Size, Str: Stars, De: Details, Sh: Shading)

**Table 2. User study results from the undergraduate group**

Combination	No of users	Probability
(Str + De)	28	0.8
(Fs + Sh + De)	23	0.657
(Str)	15	0.428
(Fs + Sh)	10	0.286
(Fs)	2	0.057
<b>Conditional probabilities</b>		
(Str+De) / ((Fs) + (Str))	0.857	
(Fs+Sh+De) / ((Fs) + (Fs+Sh))	0.75	
(Fs+Sh+de) / ((Str) + (Str+De))	0.68	
(Fs+Sh+De) / ((Str+De)+ (Fs+Sh))	0.636	
(Str + de) / (Str)	0.56	
(Str) / (Fs + Sh)	0.161	

**Table 3. Probabilities and conditional probabilities for different combinations of designs from graduate students.**

Combination	No of users	Probability
(Fs + Sh + De)	14	0.933
(Fs + Sh)	12	0.8
(Str + De)	2	0.133
(Str)	2	0.133
(Fs)	2	0.133
<b>Conditional probabilities</b>		
(Fs+Sh+De) / (not (Str+De))	1	
(Str+De) / (not (Fs+Sh+De))	1	
(Str+De) / (Str)	1	
(Fs+Sh+De) / (Fs)	1	
(Fs+Sh+De) / (Str+De)	0.5	
(Str+De) / (Fs+Sh+De)	0.071	

**Discussion:** It is clear that the undergraduate students prefer stars combined with details the most ( $p=0.8$ ), while the graduate students strongly prefer the “Font sizes + weight + shade” design ( $p=0.933$ ). To consolidate the difference, we computed conditional probabilities in order to account for the dependency of their voting, that is, whether subjects are univocal or ambivalent about the designs. For instance in the undergraduate group, the probability of liking ‘Str + De’ given the fact that they also like ‘Fs’ and ‘Str’ is quite high (0.857), indicating that many subjects have voted for these three designs simultaneously. On the other hand, the conditional probability of (Str+De) / (Fs+Sh+De) is only 0.071 in the graduate group, indicating a group more certain about their choices.

Furthermore, the difference in probability for ‘Str + De’ (0.8) and ‘Fs + Sh + De’ (0.657) in the undergraduate group is rather small, compare to their difference (0.933 vs. 0.133) in the graduate group.

An explanation for the undergraduate results (stars + detail) is that stars are linear visual forms, and therefore are the easiest to compare for their relative quantities. They are also most familiar to undergraduate students, as they have seen stars used to rate music CDs for instance. On the other hand, results from the graduates showed that if users are familiar with literature tasks and understand the difference between citation and the reputation of a music CD, then the visual signs based on their areas are more effective. The final conclusion is that we will use Fs + Sh + De as the emphasis technique for citations.

### RECOMMENDATION

We are concerned with navigational recommendation in scientific literature (implicit recommendation), which is

to provide users with additional documents based on the current document. At Citeseer, five categories of recommended documents are displayed: *cited-by* documents (those documents citing the current article), *active bibliography* (related documents at the sentence level), *similar documents based on text*, *Related documents from co-citation*, and *citations* (documents cited by the current article).

### Icon design

Two objectives were determined for the design of icons:

- 1) Easy recognition of types of recommendation based on visual signs to minimize reading
- 2) Easy recalling of the icons to facilitate task performance after a period of interruption (such as after a vacation or a period of inactivity)

Table 4 shows the two sets of icons for each of the types of recommendations used in Citeseer. Our informal user studies showed that icons from Icon 1 are slightly preferred by our users. The squares in Icon 1 set visually denote the titles being recommended, and the arrows relate these titles to the current document either in the going-out direction (citing the current document), and the coming-in direction (citing the recommended title) to the document being viewed. The icon for similar documents at sentence level shows a follow-up direction, the icon for co-cited documents shows simultaneous links from the same point (same document), and the icon for related documents based on text shows simply two documents side by side emphasizing the similarity of the whole content.

**Table 4. Two sets of icons for different categories of papers**

Category	Icon 1	Icon 2
Cited-by documents		
Cited document (active bibliography)		
Similar documents at sentence level		
Related document from co-citation		
Similar document based on text		

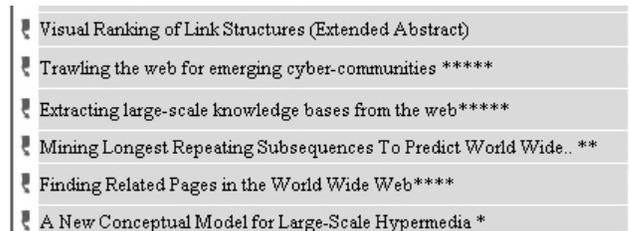
### User studies

#### Hypothesis:

Icons help users

1. to recognize the right category of information in subsequent web page visits much quicker (thus learning)

2. to quickly recall the meaning of icons after a period of task interruption (retention)



**Figure 8. Each title is marked with the same icon from the category of cited-by documents.**

#### Setup of experiment:

The users were given three design layouts to examine: the current Citeseer page layout (L1), a layout where each title is marked with an icon (L2, as shown in Figure 8) and a layout where each category of items is marked by a single icon (L3, as shown in Figure 10). Study was conducted in two steps. The first step was to determine which design is more effective for users to recognize the right category of articles. The second step was conducted after 10 days, where the same experiment was repeated. This was to determine which design enables users to more effectively retain information.

**Subject group:** There were two sets of users: users from the first set were familiar with Citeseer while the second set users were new to it.

**User tasks:** The speed of identifying types of recommended items in their respective categories.

For first time users: get familiar with Citeseer and rate the designs according to the effectiveness in recognition.

Expert users: rate the designs in terms of quickness in recognizing the category.

Returning users: rate the design in terms of effectiveness in recalling the information about a category.

**Discussion of results:** A total of 20 users were interviewed, out of which 12 were new to the Citeseer and 8 were familiar with it already. Figure 9 summarizes the user study based on their feedbacks. 16% of the new Citeseer users voted for the L1, 34% for L2, and 50% of them voted for L3. On testing the same users after 10 days, we found that 66.6% of them voted for L3, and rest opted for L2. No user voted for L1.

In the case for expert users, we observed that 37% of them voted for L1, 13% for L2, and 50% for L3. The second time of the user study, which was conducted after 10 days, showed that only 13% of the expert users voted for L1, 25% for L2, and 62% voted for L3. It is clear that the majority of users prefer the design of L3. Even though expert users initially find L1 also useful, they prefer L3 when it comes to recalling the categories.

## PUTTING IT ALL TOGETHER

Our two formal user studies showed that the non-linear mapping from citation numbers to visual values using font size, shading, and detailed information is most preferred by users to distinguish items of different reputation scales, and that icon set 1 together with layout 3 is most preferred by users to recognize and recall the type of recommended items in a visual display. We thus have designed a new interface for the NEC Citeseer (see Figure 10).

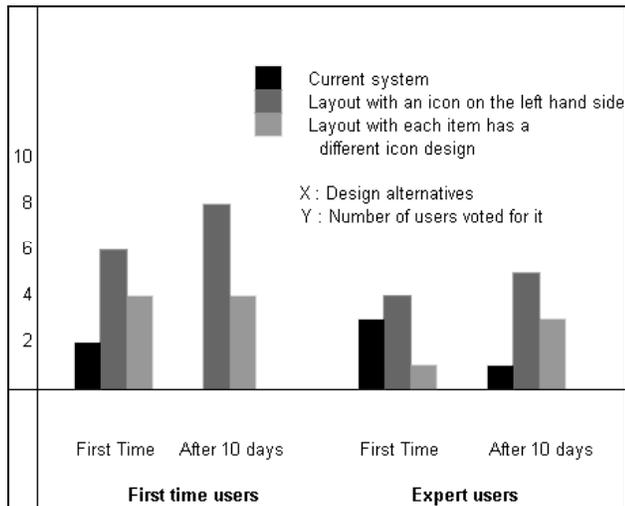


Figure 9: Results from the usability study of icon designs

## CONCLUSION

We have presented the architecture of a visually rich social environment for scientific literature search. Our goal was to investigate how such social navigation systems can augment users perception of information, as well as retention of learned signs to help them with navigation skills. We outlined our investigation strategies on various visual emphasis techniques to effectively represent reputation scales by combining the variation of font sizes, gray scales, and detail annotation. We also presented two sets of icons to represent categories of recommendations used in standard scientific digital libraries. Our user studies helped us to test and verify the most optimal emphasis techniques for reputation as well as the most effective visual signs for recommendation. We thus have achieved a compact display for a large set of, which at the same time enables users to quickly distinguish reputation among recommend information.

## ACKNOWLEDGEMENT

This work has been generously financed by the Artificial Intelligence laboratory of the Swiss Federal Institute of Technology in the framework of an exchange diploma project between EPFL and India Institute of Technology, Guwahati. We also thank the undergraduate and graduate students at EPFL for their participation in the user studies.

## REFERENCE

- [1] Basu C. and Hirsh H. (1998). *Recommendation as classification: Using social and content-based information in recommendation*. In AAAI/IAAI, pages 714--720, AAAI Press.
- [2] Bertin J. (1967/1999). *Sémiologie Graphique*, Editions de l'Ecole des Hautes Etudes en Sciences (Les reimpressions), pp. 444, Paris.
- [3] Brin S. and Page L. (1998). *The anatomy of a large-scale hypertextual Web search engine*. In Proceedings of the Seventh International World Wide Web Conference, Brisbane, Australia. Elsevier.
- [4] Card S.K. (1996). *Visualizing Retrieved Information: A Survey*. IEEE Computer Graphics and Applications, 63-67.
- [5] Card S.K., Mackinlay J.D. and Shneiderman B. (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann series in interactive technologies, Morgan Kaufmann Publishers.
- [6] Cleveland W.S. and McGill R. (1986). *An experiment in graphical perception*. Int. J. Man-Machine Studies 25, 491-500.
- [7] Dieberger A., Höök A., Svensson M., and Lönnqvist, P. (2001). *Social Navigation Research Agenda*.
- [8] Janecek P. and Pu P. (2002). *A Framework for Designing Fisheye Views to Support Multiple Semantic Contexts*. International Conference on Advanced Visual Interfaces (AVI '02), ACM Press, 2002.
- [9] Marsh S. (1994). *Formalizing Trust As a Computational Concept*. PhD thesis, University of Stirling, U.K.
- [10] Munro A., Hook K. and Benyon D. (1999). *Footprints in the snow*. In A. Munro, K. Hook, and D. Benyon, editors, *Social Navigation of Information Space*, pages 1--14. Springer.
- [11] Page L., Brin S., Motwani R. and Winograd T. (1998). *The PageRank citation ranking: Bringing order to the web*. Technical Report, Stanford University, Stanford, CA..
- [12] Resnick P. and Varian H.R., editors. (1997). *Special Issue on Recommender Systems*, volume 40. Communications of the ACM.
- [13] Sarkar M. and Brown M.H. (1992). *Graphical Fish-eye Views of Graphs*. In Proceedings of ACM Conference on Computer Human Interaction.
- [14] Yu B. and Singh M.P. (2000). *A social mechanism of reputation management in electronic communities*. In Proceedings of Fourth International Workshop on Cooperative Information Agents.

Figure 10: Final interface for CiteSeer

Cited by				Citations				
Title	Author	Year	Title	Author	Year	Title	Author	Year
<b>More</b>			<b>The Study of Instinct (151)</b>	<b>Dawkins</b>	<b>1976</b>			
Trawling the web for emerging cyber-communities (27)	Kumar, Raghavan..	1999	<b>Information retrieval (118)</b>	<b>vanRijsbergen</b>	<b>1979</b>	Journal of Experimental Psychology: Learning (4)	Anderson, Pirolli	1984
Extracting large-scale knowledge bases from the web (29)	Kumar, Raghavan..	1999	Foraging theory (1)	Stephens, Krebs	1986	Phase transitions in artificial intelligence systems (21)	Huberman, Hogg	1987
Mining Longest Repeating Subsequences To Predict World Wide.. (11)	Pitkow, Pirolli	1999	Human memory: An adaptive perspective (4)	Anderson, Milson	1989	<b>The adaptive character of thought (50)</b>	<b>Anderson</b>	<b>1990</b>
<b>Finding Related Pages in the World Wide Web (25)</b>	<b>Dean, Henzinger</b>	<b>1999</b>	Identifying appropriate in hypertext structure (2)	Botafogo, Schweideman	1991	An Object-Oriented Architecture for Text Retrieval (16)	Cutting, Pedersen et al.	1991
A New Conceptual Model for Large-Scale Hypertexts	Skillicorn	1999	Evolutionary ecology and human behavior (2)	Smith	1992	<b>Information visualization using 3D interactive animation (59)</b>	<b>Robertson, Card et al</b>	<b>1993</b>
Collective Intelligence and its Implementation on the... (1)	Francis Heylighen Center	1999	<b>Rules of the mind (82)</b>	<b>Anderson</b>	<b>1993</b>	Indexing the Internet (3)	Taubes	1995
Constructing, organizing, and visualizing collections of (2)	Terveen, Hill, Amento	1999	Results from the Third WWW Survey (2)	Pitkow, Johov	1995	<b>Information foraging in information access environments (22)</b>	<b>Pirolli, Card</b>	<b>1995</b>
Topic Shop: Enhanced Support for Browsing and Organizing	Erin Amento Loren	1999	<b>Visualizing complex hypermedia networks through multiple hie..(40)</b>	<b>Mukherjea, Foley et al</b>	<b>1995</b>	<b>Characterizing browsing strategies in the World-Wide Web (71)</b>	<b>Catledge, Pitkow</b>	<b>1995</b>
<b>An Empirical Evaluation of User Interfaces for Topic (9)</b>	<b>Amento, Hill</b>	<b>1999</b>	Gather browsing communicates the topic structure of a very (17)	Pirolli, Schank et al.	1996	<b>The WebBook and the Web Forager: An Information Workspace (59)</b>	<b>Card, Robertson et al.</b>	<b>1996</b>
Discovery of Interesting Usage Patterns from Web Data (3)	Robert Cooley Pang-Hing	1999						
A Framework for Information Visualization Spreadsheets (2)	Chi	1999						
<b>Data Preparation for Mining World Wide Web Browsing Patterns (47)</b>	<b>Robert Cooley Bamshad</b>	<b>1999</b>						
Biologically Motivated Distributed Design for Adaptive..	Rocha, Bollen	2000						
The Scent of a Site: A System for Analyzing and Predicting. (6)	Ed Chi Peter	2000						
Selective Markov Models for Predicting Web-Page Accesses (4)	Deshpande, Karpis	2000						
Information Retrieval on the Web (5)	Kobayashi, Takeda	2000						
<b>Learning to Construct Knowledge Bases from the World (29)</b>	<b>Craven, Freitag</b>	<b>2000</b>						
The Shape of the Web and its Implications for	Ek, Raghavan, Chu	2000						
Beyond Recommender Systems: Helping People Help Each Other	Terveen, Hill	2001						
Self-Organization and Identification of Web Communities	Flake, Lawrence, Giles	2002						

**Silk from a Sow's Ear: Extracting Usable Structures from the Web (1996) (Correct) (69 citations)**

Peter Pirolli, James Pitkow, Ramana Rao  
Proc. ACM Conf. Human Factors in Computing Systems, CHI

Download:  
[ntu.edu.tw/~chyun/dm/pape\\_piro96.pdf](http://ntu.edu.tw/~chyun/dm/pape_piro96.pdf)  
 Cached: [FS](#) [gz](#) [FS](#) [PDF](#) [DjVu](#) [Image](#) [Update](#) [Help](#)

From: [ntu.edu.tw/~chyun/pc](http://ntu.edu.tw/~chyun/pc) (more)  
 Home: [P.Pirolli](#) [J.Pitkow](#)  
[R.Rao](#) [\[1\]](#) [\[3\]](#) [\[4\]](#) [HPSearch](#) [\(Correct\)](#)

**NEC ResearchIndex** [Bookmark](#) [Related](#)

[\(Enter summary\)](#)

Rate this article: 1 2 3 4 5 (best)  
[Comment on this article](#)

**Abstract:** In its current implementation, the World-Wide Web lacks much of the explicit structure and strong typing found in many closed hypertext systems. While this property probably relates to the explosive acceptance of the Web, it further complicates the already difficult problem of identifying usable structures and aggregates in large hypertext collections. These reduced structures, or localities, form the basis for simplifying visualizations of and navigation through complex hypertext systems. Much of the previous research into identifying aggregates utilize graph theoretic algorithms based upon structural topology, i.e., the linkages between items. Other research has focused on. [\(Update\)](#)

of citations to this paper: [More](#)

... and structure III structure as weighted vector of page(view)s ex. [page types in Pirolli et al. B24](#) and [Cooley et al. B20](#) DP 36] [54] [15] [Info on schools indiv. school.](#) list of schools 1 parameter . 2 par.s 3 parameters Location Name. Location Name. PKDD...

... usage modeling and information capture is also relevant here. [Drawing ideas from the ACT R theory of cognition, Pirolli et al. \[45\]](#) describe how a quantitative model of information foraging can be defined. Tools for capturing history of interaction in information foraging...

BibTeX entry: [\(Update\)](#)

Pirolli96 P. Pirolli, J. Pitkow, R. Rao. Silk from a sow's ear: Extracting usable structures from the Web. Proceedings of ACM SIGCHI Conference on Human Factors in Computing, 1996. <http://citeseer.nj.nec.com/article/pirolli96silk.html> [More](#)

```

@inproceedings{ pirolli96silk,
  author = "Peter Pirolli and James Pitkow and Ramana Rao",
  title = "Silk from a Sow's Ear: Extracting Usable Structures from the Web",
  booktitle = "Proc. (ACM) Conf. Human Factors in Computing Systems, (CHI)",
  publisher = "ACM Press",
  year = "1996",
  url = "citeseer.nj.nec.com/article/pirolli96silk.html" }
    
```

Similar documents (based on text)				Related documents from Co-Citation				
Title	Author	Year	Title	Author	Year	Title	Author	Year
The Effect of Information Scent on Searching (3)	Card, Pirolli.	2000	<b>More</b>					
Distributions of Surfers' Paths through the World Wide Web.(13)	Pirolli, al	1999	Finding related pages in the World Wide Web (25)	Dean, Henzinger	1999	Data preparation for mining world wide web browsing patterns (47)	Cooley, Mobasher	1999
Exploring Browser Design Trade-offs Using a Dynamical Model (2)	Peter Pirolli	1999	<b>The anatomy of a large-scale hypertextual Web search engine (185)</b>	<b>Brin,Page</b>	<b>1998</b>	Automatic resource compilation by analyzing hyperlink structure and associated... (9)	Chakrabarti, Dom et al	1998
Mining Longest Repeating Subsequences To Predict World Wide (13)	Pitkow, Pirolli	1999	Improved algorithms for topic distillation in hyperlinked environments (63)	Bharat, Henzinger	1998	Inferring Web communities from link	Kleinberg,	1998
Web Information Retrieval (0)	Henzinger							

Similar documents (at sentence level)			
Title	Author	Year	
<b>Silk from a Sow's Ear: Extracting Usable Structures (69)</b>	<b>Pirolli, Pitkow</b>	<b>1996</b>	