Enhancing a Digital Book with a Reading Recommender

Allison Woodruff, Rich Gossweiler, James Pitkow, Ed H. Chi, Stuart K. Card

Xerox Palo Alto Research Center

3333 Coyote Hill Road

Palo Alto, CA 94304 USA

+1 650 812 4429

{woodruff, rcg, pitkow, echi, card}@parc.xerox.com

ABSTRACT

Digital books can significantly enhance the reading experience, providing many functions not available in printed books. In this paper we study a particular augmentation of digital books that provides readers with customized recommendations. We systematically explore the application of spreading activation over text and citation data to generate useful recommendations. Our findings reveal that for the tasks performed in our corpus, spreading activation over text is more useful than citation data. Further, fusing text and citation data via spreading activation results in the most useful recommendations. The fused spreading activation techniques outperform traditional text-based retrieval methods. Finally, we introduce a preliminary user interface for the display of recommendations from these algorithms.

Keywords

Spreading activation, bibliometrics, recommendations, information visualization, 3D book, degree of interest

INTRODUCTION

The digital nature of online books enables various enhancements to the reading experience that are not afforded by printed books. For example, the reader of a digital book can easily locate all occurrences of a given keyword. A primary advantage of the digital book as compared to its printed cousin is that it can be customized to suit the interests of a particular reader.

In this paper, we restrict ourselves to one enhancement of digital books—providing readers with personalized recommendations. Consider the case of an edited collection of academic papers. The reader does not always wish to read these papers in the order in which they appear in the book. Further, advances in information technology make it possible to construct an information environment in which the reader has immediate access to all of the literature referenced in a digital book. However, the related literature is often extensive, and the reader is ill

equipped to choose which documents to read next, as the knowledge required to make sensible decisions is precisely the content of the unfamiliar corpus. Readers can be aided by algorithms that attempt to predict users' changing *degree of interest* [1] in information space and by user interfaces that use these predictions to direct user attention in visualizations [2] of *information scent* [3].

To address this problem, we study the application of text- and citation-based spreading activation algorithms to the reading recommendation problem. Spreading activation is a mathematical technique for determining the relatedness of items based on their degree of association [4] and has certain properties we felt were well suited to this problem. We make several contributions:

- We present a model for recommendation that uses documents rather than terms as inputs, i.e., the inputs are the set of documents the reader has read instead of user-specified keywords. This approach reduces the reader's burden and allows us to take advantage of the extensive information available about the document.

- We present novel results in the application of spreading activation to text. The state-of-the-art in spreading activation has advanced significantly since the last (negative) results in this area [5] and we have found that the issue is worth revisiting.

- We introduce novel algorithms for the fusion of text and citation data.

- We provide a systematic evaluation of different recommendation strategies and the effects of using various data.

Our findings reveal that for the tasks performed in our corpus, spreading activation over text is more useful than over citation data. However, the fusion of text and citation data through spreading activation proves to be the most effective technique. Moreover, the new fused spreading activation text-citation technique outperforms traditional text-based retrieval methods.

We explore the utility of a reading recommender by constructing a corpus of the complete text of nearly all the documents contained in or cited by a printed book, *Readings in Information Visualization: Using Vision to Think* by Card, Mackinlay, and Shneiderman [2], hereafter referred to as RIV.

In the remainder of this section, we discuss related work. In the next section, we discuss our approach. In subsequent sections, we describe our algorithms, discuss the evaluation, and introduce a preliminary interface for presenting recommendations to the reader.

Related Work

Tracing the history of related work is a complicated task given that our methods combine three rather disparate fields. Given this and the space limitations, we focus only on attempts that fuse a) citation methods, b) text and citation methods, and c) spreading activation with either text or citations. We further restrict ourselves to examining only related research that uses citation data, leaving out similar work currently being done with the WWW, since citations are in many ways quite different from hypertext links, e.g., a primary function of links is navigation.

In what may be the first attempt to fuse citation and text representations, Salton [6] in 1963 demonstrated that citation and term data could be integrated effectively into a vector space model. Eight years later, using the SMART retrieval system, Salton [7] showed that using citations plus index terms to represent documents and queries resulted in better results than using index terms alone.

A year later in 1972, Robert Amsler [8] proposed what may be the first attempt to fuse bibliographic coupling and cocitation measures (defined in the following sections) to determine subject similarity between document pairs. Interestingly, Amsler's work predates the use of cocitation as a standalone measure of topic and document similarity. The work did not integrate textual data into the measure of document similarity. This pioneering work later inspired Bichteler and Eaton 1980 [9] who showed that re-ranking query results by the combined use of bibliographic coupling and cocitation techniques improved precision over using just bibliographic strengths. They noticed considerable variations in performance between queries. As we shall see, our work confirms this variation and provides insight into why it occurs.

Cohen and Kjeldsen 1987 [10] used a constrained spreading activation network over a knowledge base of topics to show enhanced precision and recall over keyword text methods. The knowledge base and activation networks were constructed manually. There have been several attempts to refine term and query expansion methods by various flavors of spreading activation techniques. In what appears to have placed a cap on that line of research, Salton and Buckley 1988 [5] showed that vector models performed better than several spreading activation methods using term by document matrices. While the study evaluated various normalization schemes for the spreading activation models, the spreading activation models were quite simple. As we

shall see, our work is noteworthy in that it challenges this study's findings by showing the added benefits of using spreading activation over text methods.

The methods proposed in this paper differ significantly from the prior work. First, the spreading activation model used in this paper is more sophisticated and general than earlier models. These earlier models were limited by theoretical and computational constraints. Second, most prior work focused on retrieving a set of relevant documents given a particular set of terms as the query. The methods tested in this paper deal with finding relevant documents given a set of one or more documents. As a result, the prior methods tend to use term by document representations, whereas the methods investigated in this paper use document by document matrices. Third, while a few efforts have attempted to fuse various citation data, none have attempted to combine citations, bibliographic coupling strengths, and cocitation strengths, let alone all of these with text. Finally, we are unaware of any prior work that provides a systematic investigation of the fusion of citation and text data using spreading activation.

APPROACH

The purpose of our system is to make personalized recommendations about what documents to read next. We make these recommendations using only the citation and text data from the corpus and a set of input documents of interest. We assume scenarios in which readers take different paths through the set of collected articles and have individualized interests within the field of information visualization. Traditionally, readers confronted with these choices simply make "best-guess" decisions about what to read next. These decisions are often inefficient, wasting the reader's time and energy. Alternatively, a small number seek customized recommendations from an expert on the subject, requesting the recommendations on what to read next in person, by phone, or by email. While rich in interaction and nuance, this form of interaction does not scale gracefully to handle hundreds or thousands of readers with a limited number of experts. The critical question we seek to explore is how best to provide readers with relevant individualized recommendations without having to constantly pester the experts.

In the remainder of this section, we describe the corpus and alternative approaches. This section is followed by a discussion of the algorithms used to simulate expert recommendations.

Corpus

Our study examines recommendations in the context of RIV. The printed book contains 43 articles and 9 chapter commentaries totaling 686 pages. By including all documents to which any of these articles or commentaries

refer, the total expanded RIV document count rises to 719. This expanded document, including its citation data, will be abbreviated by RIV* to avoid confusion.

We acquired and extracted text from 653 of the 719 documents, scanning and applying optical character recognition (OCR) software to the large fraction of the documents that were only available in printed form. Textual proxies (usually tables of contents, bibliographies and indices) represent some books and dissertations. The intent is for the proxies to contain most of the appropriate key words and terms. The text portion of RIV* contains some 5 million words (12,000 pages). The RIV* citation data consists of 1151 references from documents in the printed book to the 719 RIV* documents. Citations in documents not in the printed book are not considered.

Alternative Approaches
Several alternative recommendation approaches are worth considering. First, there are manual approaches. For example, one could imagine asking experts to define a fixed set of reading paths through the book and use these manually constructed paths in a digital book. This approach has several limitations. Foremost, it is limited by the predefinition of reading paths, which may or may not correspond to the actual paths of users through the book. An alternative approach would be to provide recommendations for every possible path through the book. Unfortunately, the number of paths scales as the factorial of the number of articles in the book. For the RIV book of 43 articles, the total number of unordered paths is $6.04*10^{52}$, a completely unreasonable number of recommendations to generate manually.

Second, one could use social filtering, where the usage patterns of other readers through the book are harvested to make recommendations. In today's digital and networked world, one could imagine the book's usage information collected via the Internet, recommendations generated on demand and shipped to the reader requesting a recommendation. While such a system would be dynamic and evolve over time, some form of bootstrapping would be required to seed the recommendations. That is, the usage data required to form the recommendations can not be known until some readers actually use the system to generate recommendations. Such social filtering systems also raise complicated issues such as data privacy, incentives, and biasing/spamming. Given the resource intensive nature of manual solutions and the bootstrapping and complexity issues of social filtering, we sought to systematically explore methods that would rely only upon the information contained in the RIV* book.

RECOMMENDATION ALGORITHMS
To leverage these data, we introduce a set of new methods similar to our prior work with hyperlinks and the WWW of combining citation and text data with spreading activation. We provide background material in citation analysis and spreading activation algorithms. We then describe the algorithms used in our study, including several novel algorithms.

Citation Analysis
We now define the key concepts of citation analysis, cocitation analysis and bibliographic coupling. An intuitive treatment of each will be presented, as well as formal definitions in terms of linear algebra.

Dating back to the use of the 1873 Shepard's Citations in the legal community, citation indexing has been used to harness the decisions made by authors to include references to relevant previously recorded information. Within the scientific community, these references tend to identify prior research whose methods, equipment, results, etc. influenced the current work. By capturing the semantic judgement of authors and the works of others, citation indexes create a powerful tool that serves three main application areas [11]: 1) information search and retrieval, 2) qualitative and quantitative evaluations of scientists, publications, and scientific institutions, and 3) modeling of the historical development of science.

A citation index can be represented as a directed graph (*citation network*) as well as the corresponding incidence matrix for the graph (*citation matrix*). In the former case, a directed edge between node D_i and D_j indicates that D_i references D_j and that D_j is referenced by D_i. In the latter case, the value of the cell for row D_i and column D_j denotes the number of times document D_i refers to document D_j. This number of times a document is cited is called the *citation frequency*. In this manner, the citation matrix C illustrates the "cites" relationships and the transpose of the citation matrix C^T illustrates the "is-cited-by" relationships.

The cocitation matrix and bibliographic coupling matrix (as well as a number of other interesting properties) can be readily computed from the citation matrix. If we have m source documents that contain references to n other documents with the corresponding citation matrix $C = (c_{ij})$, then

- the number of references of a given document D_i is the sum of the row vector for D_i or $(CC^T)_{ii}$;
- the number of references that documents D_i and D_j share in common (called the *bibliographic coupling strength* [12]) is given by the equation:

$$\sum_{k=1}^{n} C_{ik}C_{jk} = (CC^T)_{ij} \; ;$$

- the number of citations received by document D_i is the sum of the column vector for D_i or $(C^TC)_{ii}$;

• the number of citations which documents D_i and D_j share in common (called the *cocitation strength* [13, 14]) is given by the equation:

$$\sum_{k=1}^{m} C_{ki} C_{kj} = (C^T C)_{ij} \ ;$$

The intuition behind the value of cocitation and bibliographic coupling is as follows. Once written, the references a document D_i makes to other papers are fixed, yet additional papers can be written that reference D_i as well as cite the references in D_i. At any given point in time, one can inspect the bibliographic coupling strengths for a set of documents to gain insight into what awareness authors had of each others' work. Cocitation identifies pairs of documents that are referenced together. Frequently citing documents together implies the shared semantic judgement of authors that the pair of documents D_iD_j are related—even though the two documents may not contain a reference to each other. Cocitation strengths vary over time and can provide a glimpse into the papers that influence a particular field at any given time.

Spreading Activation

Spreading activation refers to a class of algorithms that propagate numerical values among a set of connected items. For any source of interest, activation can be spread though an association network. The highest values of the resulting activation vector represent the items most closely associated with the item of interest. Additionally, multiple sources of activation can be used to compute the interest function over several items at once. As we shall see in the next section, this feature enables the degree-of-interest function to individualize recommendations.

The particular version of spreading activation we use is the leaky capacitor model [4], which has been studied parametrically by [15]. Specifically, an activation network can be represented as a matrix R, where each element $R_{i,j}$ contains the strength of association between nodes i and j, and the diagonal contains zeros. The amount of activation that flows between nodes is determined by the activation strengths, which for our purposes correspond to bibliographic coupling and cocitation strengths. Source activation is represented by a vector C, where C_i represents the activation pumped in by node i. The dynamics of activation can be modeled over discrete steps $t = 1, 2, ...N$, with activation at step t represented by a vector $A(t)$, with element $A(t, i)$ representing the activation at node i at step t. The evolution of the flow of activation is determined by:

$$A(t) = C + MA(t-1)$$

$$M = (1 - \gamma)I + \alpha R$$

where M is a matrix that determines the flow and decay of activation among nodes, with γ determining the relaxation of node activation back to zero when it receives no

additional activation input, and α denoting the amount of activation spread from a node to its neighbors. I is the identity matrix.

The Algorithms

We introduce seven algorithms (Table 1). One (Text) is a non-spreading activation algorithm used as a baseline for comparison. The remaining algorithms are spreading activation algorithms using different association matrices.

For the comparative baseline for the various spreading activation algorithms, we used a standard text-based vector space model developed internally called Pipes. Pipes uses the standard term frequency inverse document frequency method (TFIDF) for normalization. A document by document matrix was constructed using the cosine dot product between document pairs. The resulting document similarity vectors were used to make recommendations. We refer to this non-spreading activation algorithm as the Text algorithm, and view it as the simple but traditional method with which to gauge the effectiveness of the spreading activation algorithms.

Table 1. Algorithms and association matrices.

Algorithm	*Association Matrix (R)*
Cite	C = Citation
BibCoup	BC = Bibliographic Coupling
Cocite	CC = Cocitation
Fused Citation	FC = C+BC+(3*CC)
Text	none input; outputs Document x Document
SAText	SAT = Document x Document
SATextFC	SATFC = T+(3*FC)

To leverage the citations and text data in RIV*, we created six spreading activation algorithms. The data used in each association matrix R is described in Table 1. Four methods use individual association matrices, i.e., Cite, BibCoup, Cocite, and SAText use the matrices listed in the table. Note that the document by document matrix input to SAText is the one generated by Text. The remaining two methods, Fused Citation and SAText + Fused Citation (SATextFC), use a weighted combination of the other matrices to produce their final association matrices. The weightings appear in Table 1. The weights were selected manually[1] to provide normalization across matrices. For example, with the Fused Citation method, the average cocitation strength needed to be increased by a factor of three to contribute equally with the other methods. The weighting of the matrices is supported theoretically by the additive properties of the underlying spreading activation algorithm.

---

[1] The Fisheyes scenario described below required different weightings due to abnormalities in the citation graph. We leave it to future work to develop methods to automatically correct the weights.

Introduction	*I'm a computer graphics researcher. I'm generally interested in information visualization, but I don't know much about it. I bought the book Readings in Information Visualization: Using Vision to Think, and I've read the first chapter, as listed below. What should I read next?* Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Information Visualization (Chapter 1). In S. K. Card, J. D. Mackinlay, & B. Shneiderman (Eds.), Readings in Information Visualization: Using Vision to Think. San Francisco, CA: Morgan Kaufman.
Fisheyes	*I am a VLSI chip designer. I'm writing a tool that uses fisheyes to show circuit layout. I want to read everything I can about fisheyes. I have already read Furnas' paper listed below. What should I read next?* Furnas, G. W. (1981). The FISHEYE view: a new look at structured files. Murray Hill, NJ: Bell Laboratories.
Networks	*I work at a networking company. I'm developing a tool for visualizing large networks, and I want to read about appropriate techniques. I have read the papers in Readings in Information Visualization: Using Vision to Think, Section 2.5, listed below. What should I read next?* Becker, R. A., Eick, S. G., & Wilks, A. R. (1995). Visualizing Network Data. IEEE Transactions on Visualization and Computer Graphics, 1(1 March), 16–28. Eick, S. G., & Wills, G. J. (1993, October 25-29). Navigating Large Networks with Hierarchies. Proceedings of IEEE Visualization'93 Conference, San Jose, CA, 204-210. Fairchild, K. M., Poltrock, S. E., & Furnas, G. W. (1988). SemNet: Three-dimensional representations of Large Knowledge Bases. In R. Guindon (Ed.), Cognitive Science and Its Applications for Human-Computer Interaction (pp. 201-233). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
Techniques	*I'm a researcher in the area of user interfaces. Once in a while I use an information visualization technique in my work, so I'd like to learn more about information visualization. I've read the papers listed below. What should I read next?* Bederson, B. B., & Hollan, J. D. (1994). Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics. Proceedings of UIST'94, ACM Symposium on User Interface Software and Technology, Marina del Rey, California, 17-26. Fishkin, K., & Stone, M. C. (1995, May 7-11 1995). Enhanced Dynamic Queries via Movable Filters. Proceedings of CHI'95, ACM Conference on Human Factors in Computing Systems, Denver, CO, 415-420. Furnas, G. W. (1981). The FISHEYE view: a new look at structured files. Murray Hill, NJ: Bell Laboratories. Lamping, J., & Rao, R. (1996). The Hyperbolic Browser: A Focus+Context Technique for Visualizing Large Hierarchies. Journal of Visual Languages and Computing, 7(1), 33-55. Milash, B., Plaisant, C., & Rose, A. (1996). LifeLines: Visualizing Personal Histories (Video). Conference Companion of CHI'96, ACM Conference on Human Factors in Computing Systems, Vancouver, Canada, 392-393. Plaisant, C., Rose, A., Milash, B., Widoff, S., & Shneiderman, B. (1996). LifeLines: Visualizing personal histories. Proceedings of CHI'96, ACM Conference on Human Factors in Computing Systems, Vancouver, Canada, 221-227. Rao, R., & Card, S. K. (1994). The Table Lens: Merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. Proceedings of CHI'94, ACM Conference on Human Factors in Computing Systems, Boston, MA, 318-322 and 481-482. Shneiderman, B. (1994). Dynamic queries for visual information seeking. IEEE Software, 11(6), 70-77. Spoerri, A. (1993). InfoCrystal: A visual tool for information retrieval. Proceedings of IEEE Visualization'93 Conference, San Jose, CA, 150-157.

Table 2. Scenarios used in evaluation.

EVALUATION

Experimental Design

To evaluate the effectiveness of the algorithms, we constructed four reading scenarios (Table 2), generated recommendations, and had experts rate the relevance of the recommendations. Each scenario represents a hypothetical task, and the documents listed below each scenario represent relevant documents a reader might have read up to the point of requesting a recommendation. We deliberately choose to vary both the number of documents read as well as whether the documents were from the RIV book or referenced documents to gain insight into the behavior of the algorithms under various conditions.

The recommendations for Text were generated by the following method: the row vector corresponding to the read document was sorted and the top documents selected. In cases in which more than one document was read, sorted vectors for each read document were merged into a single list from which the top ranked documents were selected.

For the spreading activation algorithms, the documents read by the hypothetical reader were used in the source activation vector to pump activation. Since the average values in the association matrices for the citation matrices were much lower than in the document by document matrix, we used an alpha of 1 for the citation methods and an alpha of 0.01 for the SAText methods. Activation was spread for ten iterations and in all cases quickly converged. The top ten values from the final activation

vector were selected as the final recommendations. As a further baseline for the comparisons, we included a random recommendation generator that generated random documents.

In order to determine the ability of each algorithm, we had three information experts rate the relevance of the recommendations. The evaluators included two authors of the RIV book and an expert in the area of information visualization. Each scenario was presented to the evaluators with a random permutation of the recommended documents. Evaluators were asked to rate each document's "usefulness" to the reader described in the scenario on a scale of 1 (least useful) to 5 (most useful). The evaluators took approximately half an hour to complete the task. The correlation between the rankings of the evaluators ranged from 0.50 to 0.59.

Analysis of Precision

To assess the quality of each recommendation, we computed the geometric mean of the experts' ratings for each document within the context of a given scenario. The mean was then compared against a threshold value to assess whether the recommendation was *useful* (geometric mean 4 or higher) or *somewhat useful* (geometric mean 3 or higher), e.g., if the geometric mean of the experts' ratings of a document was 3.9, it was a *somewhat useful* recommendation.

From the mean, we computed various precision metrics in a similar manner to [16]. Recommendations are presented in ranked lists from 1 to 10, with 1 being the highest

recommended. For a given algorithm, the precision at some ranking r is the total number of useful recommendations within the first r answers divided by r (we also calculate the precision for somewhat useful recommendations). For example, suppose an algorithm had a *not useful* first recommendation and a *useful* second recommendation. In this case, its precision at 1 would be 0 and its precision at 2 would be 1/2. Intuitively, precision tells us the percentage of *useful* versus *not useful* rankings for the first r recommendations made by the algorithm.

We next define an aggregate metric, the average precision of an algorithm. The average precision of an algorithm is the sum of its precision at all ranks divided by the total number of ranks. In our case, since each algorithm made 10 recommendations, the average precision is the sum of the precision at ranks 1 to 10 divided by 10. Intuitively, this metric assigns higher values to algorithms that get useful recommendations early in the rankings. For example, suppose algorithm A had 5 *useful* rankings followed by 5 *not useful* rankings and algorithm B had 5 *not useful* rankings followed by 5 *useful* rankings. In this case, algorithm A would have an average precision of 0.86 while algorithm B would have an average precision of 0.18.

The first seven data columns in Table 3 contain the average precision for each of the algorithms across all scenarios. The random algorithm had an average precision of zero in almost every case, so we omit it from the table. The final column compares the relative performance of SATextFC and Text (computed by dividing the average precision of SATextFC by the average precision of Text and subtracting 1).

Discussion of Average Precision

We begin with several general observations. First, we see that the average precision values vary by scenario as was noted by Salton and Buckley 1988 [5]. In particular, the algorithms performed quite differently on the Networks scenario, as we will discuss further below. Second, the average precision is much higher for the *somewhat useful* than for the *useful* conditions, as is to be expected. The relative orderings of the algorithms remain fairly stable across these conditions, with the exception of Cocite, which drops dramatically in the more stringent *useful* condition. Third, there is considerable variability in the precision of each of the individual citation methods. With the less stringent criteria, the Cocite method performs the best, but in the more stringent condition, bibliographic coupling performs the best. As one would expect, the Fused Citation method provides a smoother precision across scenarios. Finally, we see that the algorithms that use text compare favorably to those that use citation structure.

More specifically, we observe that our new algorithms SATextFC and SAText both compare favorably to Text

and all citation algorithms, particularly in the demanding *useful* condition, in which they yield a 47% improvement over Text overall. The SATextFC fusion method appears somewhat superior to SAText in almost all cases except the Network scenario; this decreases its performance so that on average these two algorithms have very similar behavior. If the network scenario is not considered, SATextFC yields on average a 71% improvement over Text in the *useful* condition and a 22% improvement over SAText.

Discussion of Average Precision by Scenario

To gain more insights into variation by task, we intentionally chose scenarios with different characteristics. Below, we make observations about the behavior of these algorithms for these particular scenarios.

The Introduction scenario is a broad, vague query. The Introduction contains a large amount of highly relevant text and a large number of citations. The citation algorithms (C, BC, CC) did generate a number of unique recommendations that were *useful* or *somewhat useful*. The text similarity algorithms did very well in this scenario, presumably because the input document was a strong indicator of the reader's interests. SATextFC did slightly better than either strategy.

Table 3. Average precision for all scenarios.

	Cite	BibCoup	Cocite	FC	Text	SAText	SATextFC	SATextFC vs Text
Somewhat Useful								
Introduction	0.52	0.46	—	0.24	0.86	0.86	0.93	9%
Fisheyes	0.49	0.80	0.50	0.64	0.87	0.84	0.97	11%
Networks	0.14	0.05	0.49	0.20	0.38	0.62	0.34	-11%
Techniques	0.70	0.58	0.91	0.84	0.72	0.73	0.95	32%
Average	*0.46*	*0.47*	*0.63*	*0.48*	*0.71*	*0.76*	*0.80*	*13%*
Useful								
Introduction	0.15	0.17	0.00	0.19	0.37	0.48	0.56	52%
Fisheyes	0.49	0.73	0.37	0.47	0.77	0.79	0.95	22%
Networks	0.00	0.00	0.00	0.01	0.36	0.59	0.26	-29%
Techniques	0.19	0.02	0.05	0.14	0.00	0.33	0.44	—
Average	*0.21*	*0.23*	*0.14*	*0.20*	*0.38*	*0.55*	*0.55*	*47%*

The Fisheyes scenario is the most focussed query. The input document contains specific terminology and is cited by many other documents in the collection. Given the strong text and citation cues, it is not surprising that citation and text algorithms both performed well. Most interesting, the fused SATextFC did better than both the citation and text measures independently, successfully integrating, which resulted in near perfect precision. For example, SATextFC ranked as its second choice a document that had received lower rankings from the text and citation measures (Cocite 5; Text 4; SAText 4).
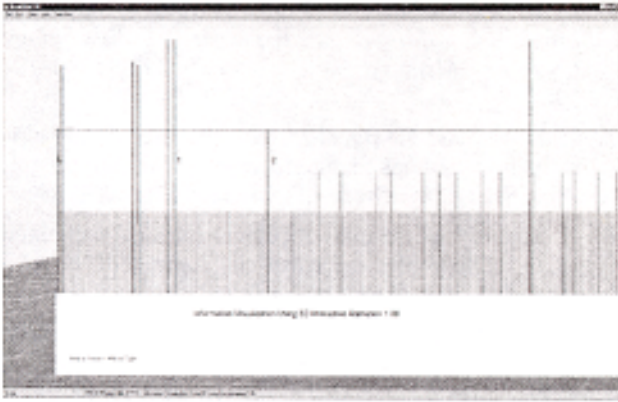
Figure 1. Close-up of a section of the Book Ruler.

The Network query is a less focussed text query. This was the most difficult scenario for all the algorithms. The citation algorithms performed particularly poorly on this scenario, which we believe results from the lack of cohesive citations in this sub-discipline. In this instance, the seminal papers do not cite the same authoritative references (low BibCoup), but are cocited by a fair number of later papers (high Cocite). We were surprised and excited to find this behavior of the citation methods. The document cosine algorithm was outperformed by the SAText algorithm by 62% for *somewhat useful* recommendations and 65% for *useful* recommendations. This is the only case in which SATextFC performed worse than the other text algorithms. This occurred because the citation recommendations were so poor that they lowered the quality of SATextFC recommendations.

The Techniques query is very broad query, using nine input documents. In the *somewhat useful* conditions, both the citation algorithms and the text algorithms did well, with SATextFC having the best performance, closely followed by Cocite. The citation methods performed their best as a result of the well-defined input citation structure. In the *useful* condition, the citation algorithms and Text have very poor performance, indicating that the overall quality of the recommendations was moderately useful. The SAText and SATextFC were able to produce higher quality recommendations than the other algorithms. We credit the improved performance to the manner in which the spreading activation algorithm reinforces useful documents during each iteration.

Comparison of Algorithms

To assess the statistical significance of our findings, we performed the Wilcoxon Sums of Ranks Tests for each pair of algorithms using a significance level of 0.01. For each algorithm, the input was an ordered list of the expert ratings for the algorithms' recommendations (recall that values for expert ratings were computed by taking the geometric mean of all experts). From this analysis, we can conclude that the difference between SATextFC and Text was statistically significant. However, the differences between SATextFC by SAText and SAText by Text were
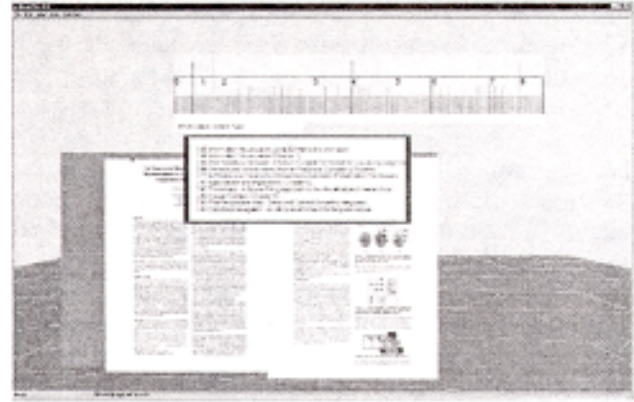

Figure 2. Book, Book Ruler, and recommendation list.

not statistically significant. All text methods were significantly better than the citation and fused citation methods.

USER INTERFACE

Having determined that the recommendation algorithms, especially SATextFC, produce reasonable recommendations, we now discuss a simple user interface prototype that uses the recommendations to augment the user's reading experience. Our prototype has a recommendation engine and two major graphical components: the 3D Book [17] and the Book Ruler. The 3D Book presents a graphical representation of RIV*. The reader can interact with the book, e.g., to turn pages. The Book Ruler uses graduated lines to give the reader an overview of the contents of the 3D Book. Each major graduation on the Book Ruler represents a chapter commentary, the next finer divisions represent articles in the corresponding chapter, and the finest graduations represent references for each article (these graduated lines appear below the horizontal line in Figure 1). Selecting a graduation turns the pages of the book to that article. Much like the Data Slider [18], the Book Ruler allows the reader to brush over the graduations. Brushing highlights the graduation and highlights any other locations where the same document appears, allowing the reader to see the pattern of references to a particular document.

By interacting with the Book Ruler or by issuing queries to the system, the reader can select a set of documents. A simple user interface gesture feeds the current selection set into the recommendation engine to indicate, "Given I've read this, what should I read next?" In response, the system performs a recommendation analysis and displays the results as taller, colored bars above the horizontal line in the Book Ruler. The reader can see where recommendations appear in the book and make quick comparisons of relative value. Results also appear in a pop-up text box (Figure 2).

3D book metaphors have been used in user interfaces for some time (for a review, see [17]). The 3D book metaphor helps orient users and maintains correspondence between physical and digital versions of documents. What we

believe is new is the integration of a 3D book on this scale with citation links and algorithms for computing the user's degree of interest in the material. BellCore's Superbook [19], for example, was designed to transform existing electronic documents into hypertext documents with indexing and a fisheye table of contents. However, Superbook did not use a simulation of a physical book, did not include the text of all its references, and did not use our algorithms based on citation analysis for degree of interest, although it did have a notion of degree of interest. Our earlier WebBook [17] could not handle the data on this scale and was not integrated with a degree of interest algorithm.

CONCLUSION

The reading experience of digital books can be enhanced in a number of ways. In this paper, we have focused on automatically generating recommendations of further reading material based on various user scenarios. In our evaluation, we compared our algorithms to standard information retrieval algorithms. We demonstrated that spreading activation over text and fusing text with citation data techniques are very effective, according to expert evaluation. While our algorithms make successful recommendations using only text and citation data, we believe we could enhance these algorithms by adding usage data, and spreading activation is a natural mechanism for fusing these three disparate data types.

The corpus we considered is in the research literature genre, where a particular paper or book often serves as an initial exploration into a group or groups of documents. We believe the results support the exploration of the new techniques in other linked information environments such as the digital libraries or the Web.

ACKNOWLEDGMENTS

REFERENCES

1. G. W. Furnas, "The FISHEYE view: a new look at structured files," Bell Laboratories Technical Report, reproduced in *Readings in Information Visualization: Using Vision to Think*, S. K. Card, J. D. Mackinlay, and B. Shneiderman, Eds. San Francisco: Morgan Kaufmann Publishers, Inc., 1981, 312-330.

2. S. K. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*. San Francisco, California: Morgan-Kaufmann, 1999.

3. P. Pirolli and S. K. Card, "Information foraging," *Psychological Review* 106(4):643-675, 1999.

4. J. R. Anderson and P. L. Pirolli, "Spread of activation," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10(4):791-798, 1984.

5. G. Salton and C. Buckley, "On the use of spreading activation methods in automatic information retrieval," Proc. SIGIR '88, Grenoble, France, 147-160, 1988.

6. G. Salton, "Associative document retrieval techniques using bibliographic information," *Journal of the ACM* 10(4):440-457, 1963.

7. G. Salton, "Automatic indexing using bibliographic citations," *Journal of Documentation* 27(2):98-110, 1971.

8. R. Amsler, "Applications of citation-based automatic classification," Linguistics Research Center, Univ. Texas at Austin, Technical Report 72-14, Dec. 1972.

9. J. Bichteler and E. A. Eaton, "The combined use of bibliographic coupling and cocitation for document retrieval," *Journal of the American Society for Information Science* 31(4):278-282, 1980.

10. P. R. Cohen and R. Kjeldsen, "Information retrieval by constrained spreading activation in semantic networks," *Information Processing and Management* 23(4):255-268, 1987.

11. P. Zunde, "Structural models of complex information sources," *Information Storage and Retrieval*, 7(1):1-18, 1971.

12. M. M. Kessler, "Bibliographic coupling between scientific papers," *American Documentation* 14:10-25, 1963.

13. I. V. Marshakova, "System of document connectionism based on references," *Nauchno-Teknicheskaya Informatsiya*, Series 2:2-6, 1973.

14. H. Small, "Co-citation in the scientific literature: a new measure of the relationship between two documents," *Journal of the American Society for Information Science* 24(4):265-269, 1973.

15. B. A. Huberman and T. Hogg, "Phase transitions in artificial intelligence systems," *Artificial Intelligence* 33(2):155-171, 1987.

16. J. Dean and M. R. Henzinger, "Finding related pages in the World-Wide Web," *Computer Networks & ISDN Systems* 31(11-16):1467-1479, 1999.

17. S. K. Card, G. G. Robertson, and W. York, "The WebBook and the Web Forager: an information workspace for the World-Wide Web," Proc. SIGCHI '96, 111-117, Vancouver, BC, Canada, 1996.

18. S. G. Eick, "Data visualization sliders," Proc. ACM UIST '94, 119-120, Marina del Rey, CA, 1994.

19. J. R. Remde, L. M. Gomez, and T. K. Landauer, "Superbook: an automatic tool for information exploration," Proc. ACM Hypertext '87 Conference, 187-188, Chapel Hill, NC 1987.

Enhancing a Digital Book with a Reading Recommender (Page 153)

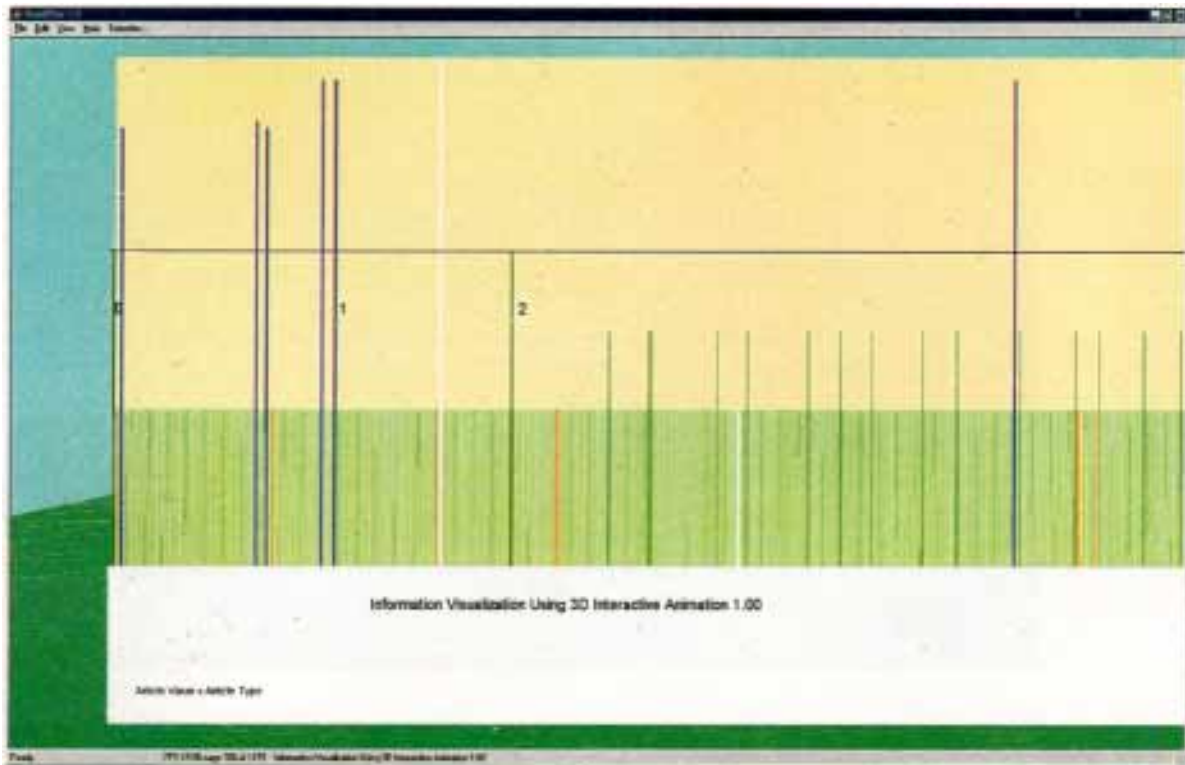Allison Woodruff, Rich Gossweiler, James Pitkow, Ed H. Chi, Stuart K. Card
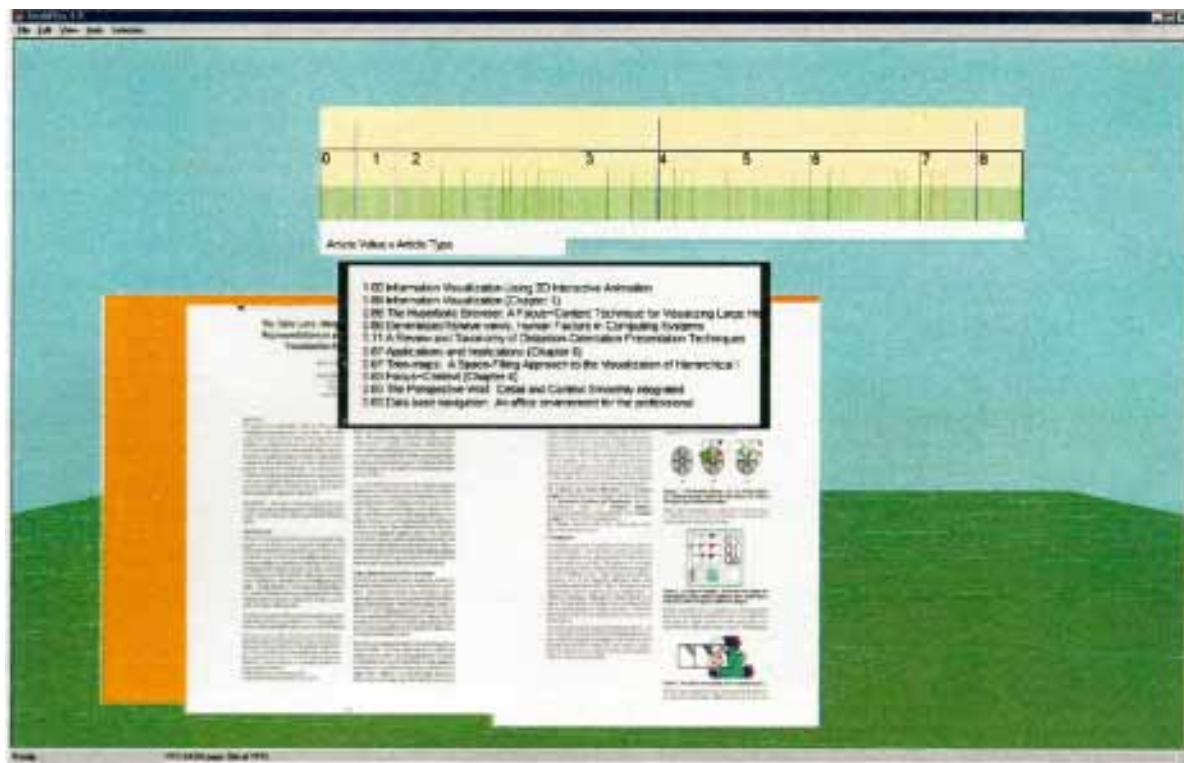

Figure 3. Close-up of a section of the Book Ruler.

Figure 4. Book, Book Ruler, and recommendation list.