**Independent Study Report**
Rosta Farzan
Intelligent System Program
Fall 2004

# Improving Social Navigation Support Using Time Spent Reading

## 1 Introduction

KnowledgeSea is a mixed corpus C programming resource that tries to bridge the gap between open (unlimited resources) and closed corpus (limited resources). Closed corpus materials are in the form of lecture notes that are well designed for the purpose of the course. Open-corpus materials are the set of the links to online resources for C programming. KnowledgeSea is designed to help users navigate from lectures to relevant online tutorials in a map-based horizontal navigation format. The map is self-organized using neural network techniques to cluster similar documents together [2]. Every cell of the KnowledgeSea map includes links to online material that are related to keyword presented on the cell. The adjacent cells present similar material. To facilitate the student navigation, KnowledgeSea provides traffic and annotation based social navigation support. In the new phase of the project as a work for my independent study during Fall 2004, I have been looking into improving traffic based navigation support by taking into account time spent reading each tutorial page by each students.

### 1.1 Traffic based social navigation support

Traffic based navigation support is offered in the KnowledgeSea map through coloring schema. Each cell on the map has different background. The color of the background represents the group traffic that is the indicator of the number of visits group of students has made on the documents inside each cell. All the cells of the map initially have very bright shade of blue. As students start visiting the pages, the background color of the cell including visited documents gets darker and darker. The changes of the background are in a very smooth way that create large spectrum of blue in the map. In this way, student could easily follow footprint of others by visiting cells with the darker background. In addition, there is a small “man icon” on each cell on the map. The color of this icon represents the traffic made by the specific student that means the number of visits she or he has made on documents inside the specific cell. The same schema is applied here as well; the color of the icon is set to very bright blue and as the student start visiting more and more pages the color of the icons gets darker. This makes the map personalized for each student. In this way students could easily follow others and also compare their navigation behavior with other students. If the color of the person icon is brighter than the background it means that the specific student has made fewer visits than the average number of visits in her or his group. Therefore this could give the student a clue for what page to visit next.

When students choose a specific cell, they get to see the details about the document inside each cell. Similar to the map, the coloring schema is being used to provide social navigation support inside cell content pages. Here, all the links inside the cell content interface are annotated with group and student traffic. Correspondingly, the background color represents the number of visits made by group of students and the icon color represents the number of visits made by the specific student.

Once students click on a link inside the cell content window, the actual tutorial page is being opened in a new window. All the links inside pages are also annotated to present the traffic based social navigation support in the same manner.

### 1.2 Annotation based social navigation support

KnowledgeSea system offers the ability to annotate tutorial pages to the students. Students could benefit from writing and annotating in two important ways: First, writing is proven to be a meta-cognitive strategy that involves students in active learning and engages them with more effectively in the learning process. As summarized by Bonifazi et al, [1] annotation could help learners to remember better by highlighting the most significant part of a text, could encourage learners to think when they add more ideas to what they are reading, and could help learners to clarify and make sense of the material while they try to reshape the information [1]. Second, students' annotations could be used to build stronger evidence of interest in read document that could improve social navigation support. Students' annotations could create an important trail for other learners to follow.

To provide the ability to write notes, all pages of AnnotatEd include an annotation frame. The annotation ability is designed in a very simple and clear way to avoid any extraneous load on students. The annotation can be in the form of writing free notes or highlighting. The written note can be private or public and can be in the form general, praise, or problem. At the top of the annotation section students can see all previously written notes.

Annotation based social navigation support is offered by augmenting the resources with a sticky note icon inside a colored square. This icon represents both the students' own and group annotation density. The background becomes a darker shade of yellow with increasing group annotation. The central "sticky note" portion of the icon becomes darker with increasing density of the students' own annotation.

Refer to [3] for more details on the KnowledgeSea system.

# 2 Improving traffic based navigation support using TSR

As described above previously traffic based social navigation support was offered based on the number of visits made to the page. However, just visiting a page does not necessarily mean interest in the page. Looking into the literature, there are several researches in finding out better and stronger indicators of implicit feedback. Miller et al in [6] show that there is a high correlation between users' rating of news article and time spent reading the article. They also show that there is no correlation between the length of article and time spent reading the article. Morita and Shinoda in [7] presents the same result and show a very high correlation between users' interest (specified through explicit rating) and time spent reading. Rafter and Smyth describe the same result in the job search domain. They present high correlation between users' interest in a job and the time they spent reading the job description [8]. Claypool et al in [4] talk about importance of using implicit indicator of feedback in recommender system and they describe several problems associated with using explicit feedback such as user tendency in reading a lot more than providing any feedback or issue of stopping users regular process for providing explicit feedback. They examine different types of implicit feedback such as time spent reading, number of mouse click, time spent moving the mouse, and time spent scrolling. They found time spent reading (TSR) a page as one of the most important implicit indicator of interest. Therefore the previous research nicely supports our goal of adding TSR information into our traffic based navigation support.
As a result, the color of the background and the man icon – described in the previous section - would be changed based on the time each student is spending on reading a page and not just visiting a page. In the previous version, each visit would increase the visited counter by 1 disregarding the TSR. Therefore, the main challenge is increasing the visited count by a meaningful number based on TSR. The key point in finding the increasing number is setting a reading threshold. Obviously any TSR above the threshold will increase the visited count by 1, but when the TSR is below the threshold the visited count should be increases by a number less than one. Although in our early discussion we considered that TSR will be affected by the length of the article and the reading threshold should have some reflection of the length of the article, previous research changed our direction at the beginning. Previous research shows that there is very low correlation between the length of an article and TSR of that article. In reality readers would ignore reading an article very quickly if it does not seem interesting no matter what the length of the article is [4], [6]. Following the recommendation of the literature of recommender systems, I considered the median TSR for each page as the reading threshold. To calculate the threshold I employed our previous data and calculated the TSR for each document for each student over each access and considered the mean of these values as the reading threshold. However after looking into the data I figured out this approach does not suit our need perfectly. Knowledge Sea currently includes over 20000 online documents and the number of documents

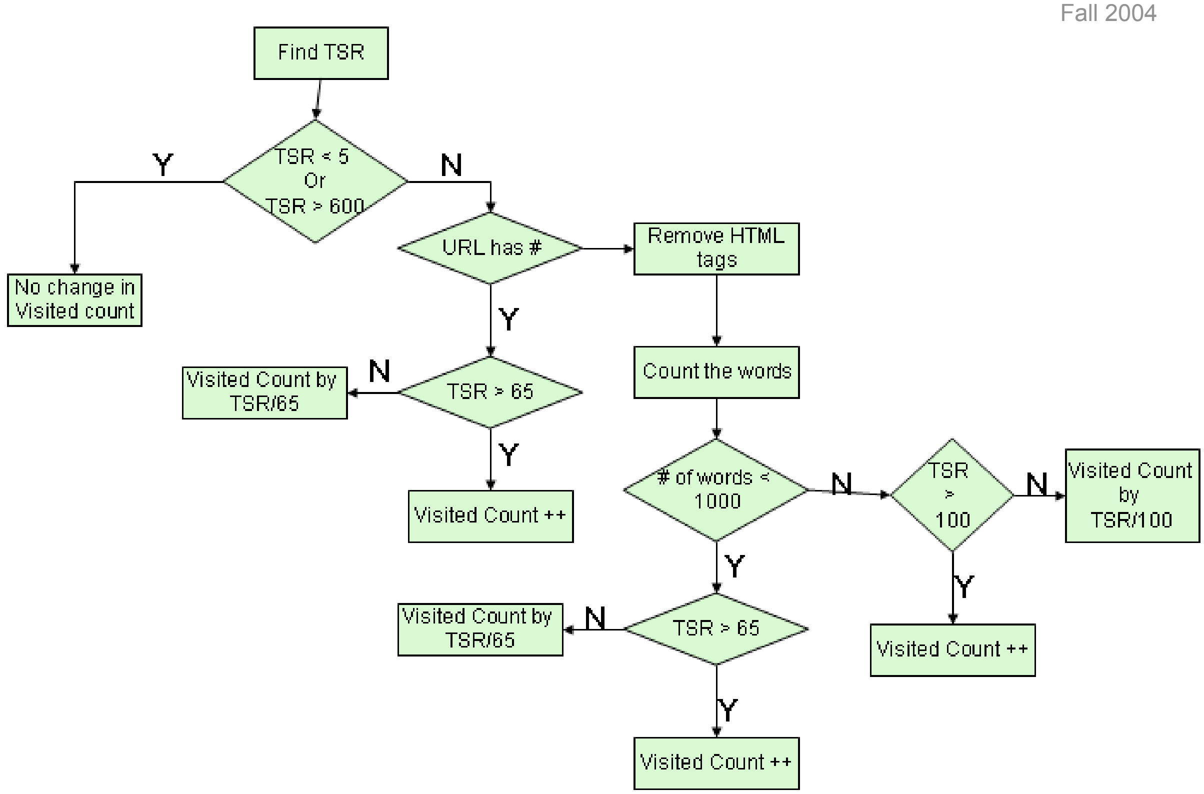
are growing as learners continue to use the system. However, the number of students who have been using the system is very limited. Therefore the chance of reading the same page with many users is very low which means many pages have had very few visits. As a result the reading threshold in many cases will be biased with very few numbers of visits. For example a page could have only one visit with a very short TSR which will set the threshold too low and another page could have only a long TSR which will set the threshold too high. Using the median approach would end up having very high variances threshold for different pages without a meaningful reason.

As a result I decided to find a common threshold for all documents. At this point we decided to examine our first hypothesis of effect of page length in TSR based on previous data. Especially because some programming tutorial pages which describe a brief concept have very short length and a page might be interested for a student and read in a short time just because the page is very short.

The annotation ability of the system could help us to evaluate our hypothesis of relationship between time spent reading a page and the length of the page. To evaluate our hypothesis we looked at time spent reading each annotated page by the annotator student. We looked at the data for each semester separately to exclude effect of other parameters. In spring 2004, overall 69 documents were annotated and our analysis shows a significant difference in TSR for pages with less than 1000 words and more than 1000 words ($\alpha$=0.1, t-test). In Fall 2004, overall 20 documents were annotated and again our analysis shows a significant difference in TSR for pages with less and more than 1000 words ($\alpha$=0.1, t-test). Therefore, we concluded that length of the page should be considered in the algorithm for updating the visited count.

Next step is figuring out the value to be added to visited count based on TSR and length of the page. To specify this value again I made use of our previous data. The average time spent reading annotated pages by annotator with less than 1000 words for the previous data is about 65 seconds and the average TSR for documents with more than 1000 words is about 100 seconds.

Taking into account abovementioned factors, the following flowchart describes the algorithm to updated visited count. As shown in the flowchart, first step we try to discard noisy data by ignoring pages with TSR less than 5 seconds or greater than 10 minutes. Moreover, we had to take into account pages with several sections. When a page has more than once section it is not clear which part is the focus for the student at each access time. Therefore, the length of the page is not clear. In these cases we consider the page as a short page (pages with less than 1000 words) to be on the safe side. The pages with several sections are being determined by having # in the URL. Consequently, for each access to a short page the visited count is increased by 1 if TSR is greater than 65 seconds and is increased by TSR/65 if TSR is less than 65 seconds. Similarly, for long pages (pages with more than 1000 words) the visited count is increased by 1 if TSR is greater than 100 seconds and is increases by TSR/100 if TSR is less than 100 seconds. We understand that this is pretty simple algorithm since it does not take into account individual different speed of reading. However, we decided to evaluate the simple algorithm before spending more time developing more elaborated algorithm.




# 3 Evaluation

My early evaluation presents some promising result in adding TSR information into our traffic based navigation support. For the primary evaluation, we have contrasted the simple click based navigation support with TSR based navigation support by comparing the difference of average click on important pages with average TSR based click on important pages. We consider pages with students' annotation as important pages for this stage of evaluation. The following table shows that the TSR Click (which is the average click normalized based on TSR following the above algorithm) is better differentiator of important pages from not important pages.

| | **Num** | **Mean Click** | **TSR Click** |
|---|---|---|---|
| **Annotated** | 38 | 4.87 | 2.55 |
| **Not Annotated** | 339 | 1.88 | 0.71 |
| **p-value** | | .00005 | .000002 |

In the next phase of the project I plan to perform more evaluation to assess the effect of adding TSR to the system. I will evaluate the new features both by using the previous data to and by trying the new system with new subjects to assess the following hypothesis:

1. The background color for pages with larger number of visits but short TSR is lighter than pages with less number of visits but longer TSR (Test the accuracy of the algorithm)
2. The annotated pages will have darker background (Test the accuracy of the algorithm and reveal the correlation of TSR and interest – done partially at this stage)

3. The important pages (identified by the instructor) among all visited pages by students will have darker background.
4. The average search time will be shorter. To evaluate this I will look at the length of the path till spending significant amount of time reading a page. I will compare the length of the path in the systems with and without consideration of TSR.

# References

[1] Bonifazi F., Levialdi S., Rizzo P., Trinchese R. (2002), UCA: A web-based annotation tool supporting e-learning. Proceedings of the Working Conference on Advanced Visual Interfaces: AVI 2002. ACM Press, New York (NY), pp. 123-128.

[2] Brusilovsky P. and Rizzo R.(2002), Map-Based Horizontal Navigation in Educational Hypertext. In Proceedings of Hypertext 2002, University Of Maryland, College Park, USA

[3] Brusilovsky P., Chavan, G, Farzan, R. (2004), Social adaptive navigation support for open corpus electronic textbooks. In Proceedings of Adaptive Hypermedia 2004, Eindhoven University of Technology, the Netherlands.

[4] Claypool, M., Le, P., Waseda, M., and Brown, D. (2001). Implicit interest indicators. In Proceedings of ACM Intelligent User Interfaces Conference (IUI), (Winner! Best paper award)

[5] Dron J., Boyne C., and Mitchell R. (2001). Footpaths in the stuff swamp. In: Fowler, W. and Hasebrook, J. (eds.) Proc. of WebNet'2001, World Conference of the WWW and Internet, Orlando, FL, AACE (2001) 323-328

[6] Miller B., Riedl J., and Konstan J. "GroupLens for Usenet: Experiences in Applying Collaborative Filtering to a Social Information System" in C. Leug and D. Fisher (eds.) From Usenet to CoWebs: Interacting with Social Information Spaces. Springer-Verlag, 2002

[7] Morita, M., Shinoda, Y. (1994): Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. Proceedings of the 17th Annual International SIGIR Conference on Research and Development. pp. 272—281

[8] Rafter R., Smyth, B. (2001) Passive Profiling from Server Logs in an Online Recruitment Environment. In Proceedings of the IJCAI Workshop on Intelligent Techniques for Web Personalization (ITWP 2001) Seattle, Washington, USA