Investigating Users' Needs and Behaviors for Social Search

Jae-wook Ahn and Peter Brusilovsky

Personalized & Adaptive Web Systems Group (PAWS)
Department of Information Science and Telecommunications
School of Information Sciences
University of Pittsburgh
http://www.sis.pitt.edu/~paws

Introduction

- Web information explosion – Web search
- Problems of traditional Web search
  - Rank by similarity between query and documents
  - Mismatch between query & document space
    - Ambiguous & short user queries
    - Not in the same space
  - “One size fits all” approach – no personalization
- New approaches
  - PageRank – re-rank by incoming link counts (general popularity)
  - Community-based search – re-rank by community relevance
- KnowledgeSea “Social Search”
  - Supports social guidance in search context in KnowledgeSea II

Social Navigation

- Use past users' interaction with the system to support information navigation
  1. Support a known social phenomenon
     - We follow similar people's "footprints"
  2. Self organization
     - Function without human's manual endeavors
- Examples
  - Browsing: classic SNS ("footprints" system)
  - Recommendation: collaborative filtering (MovieLens Amazon recommender system)
  - Ad-hoc Search: Social Search in I-SPY and KSII

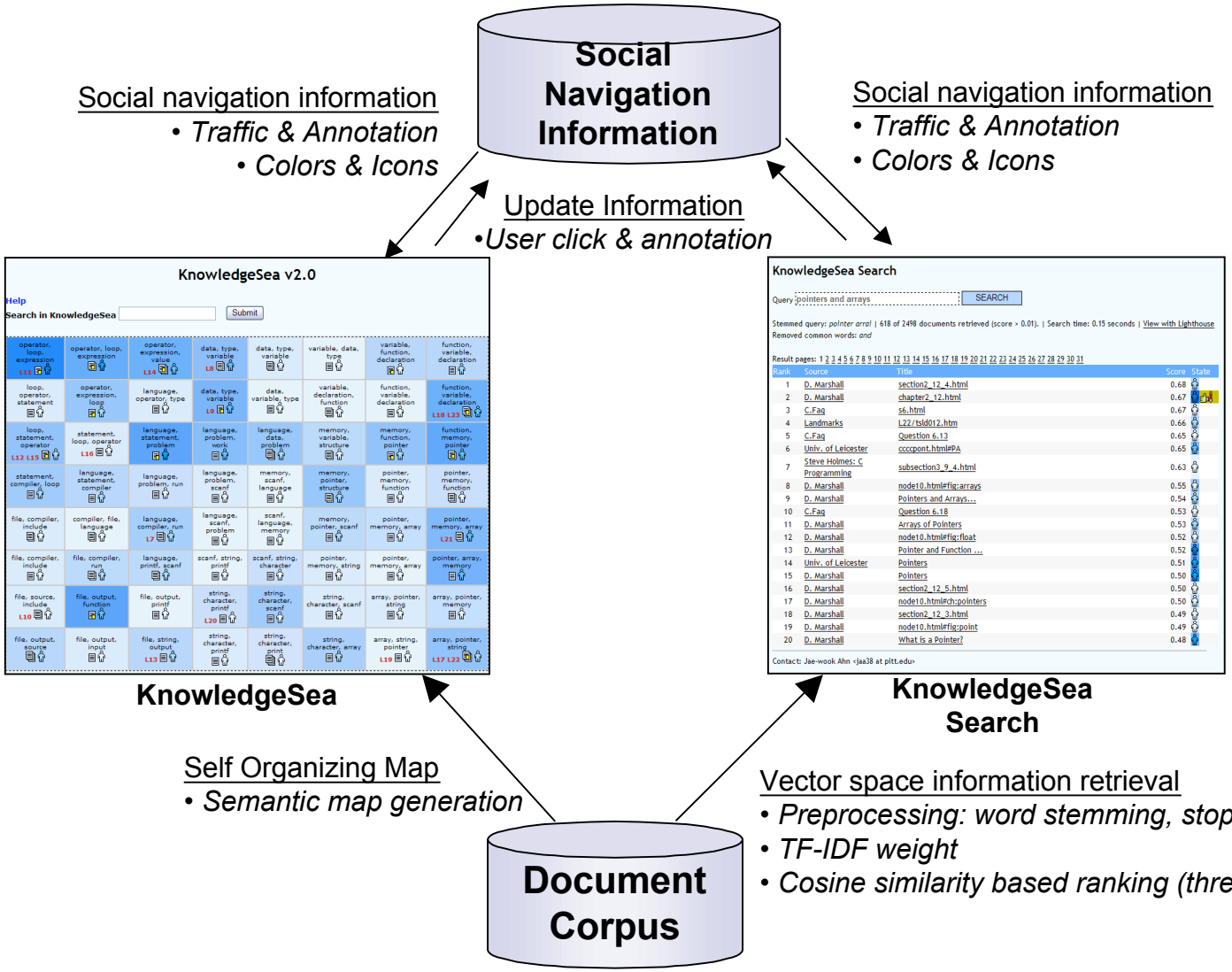
Social Search

- I-SPY (Smyth, et al.)
  - Relies on search histories of similar users
    - Repetition of query terms should be high
  - Re-rank by community and individual query based popularity
  - Query-document frequency matrix
- Knowledge Sea II
  - Relies on link selection and page comments/highlights
  - Adaptively annotate search results
  - User action history

Knowledge Sea II

- Web based social navigation support system
- Includes open corpus – uses self organization
- Information Architecture
  - Several hierachical electronic hypertextbooks
  - Knowledge Map build with SOM (Self Organizing Map)
- Social navigation support
  - Color and Icons
  - Different level of traffic and annotations
  - Different types of annotations

Knowledge See II+ Architecture



System Design and Implementation

- Document Corpus
  - C language tutorials and slides
  - Shares document URL DB with KnowledgeSea
  - Fetch and index documents → searchable
- Stemming
  - Porter's algorithm
- Stopwords
  - Terms less contributing for document discrimination
  - C keywords
    - Overlaps with some stopword (for, if, while, etc.)

System Design and Implementation (cont.)

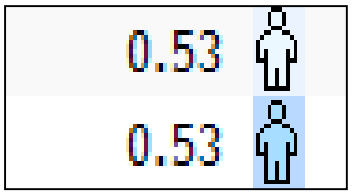
- Term weighting
  - TF-IDF
    - TF (Term Frequency)
      - importance of a term in a document
    - IDF (Inverse Document Frequency)
      - concentration of a term
      - importance of a document given a term
    - TF * IDF as term weights
- Retrieval model
  - Vector space (Salton)
    - Documents and queries are represented as vector of terms
    - Document vector components – TF-IDF weights
    - Query vector components – Binary
  - Rank by cosine similarity

Presentation of Search Results

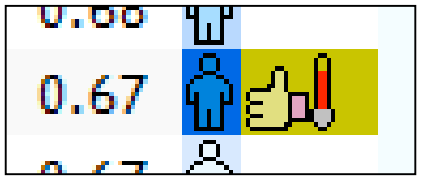
- Conventional ranked presentation of search results
  - Rank, Document source, Title, Relevance Score (similarity)
- Social navigation visual cues for each link
  - Traffic-based
    - *How many times users clicked (selected and viewed) the page behind the link*
  - Annotation-based
    - *Annotations/highlighs made by users to this page*

Visual Cues

- Traffic-based
  - More group traffic
    - Darker background color
  - More user traffic than others
    - Darker foreground color of the “human” icon
  - 0~9 traffic levels
- Annotation-based
  - More group annotations
    - Darker background color
  - User own annotation
    - Foreground color
    - Type: Sticky notes, Thumbs-up, Question
  - General attitude
    - Page “quality” temperature

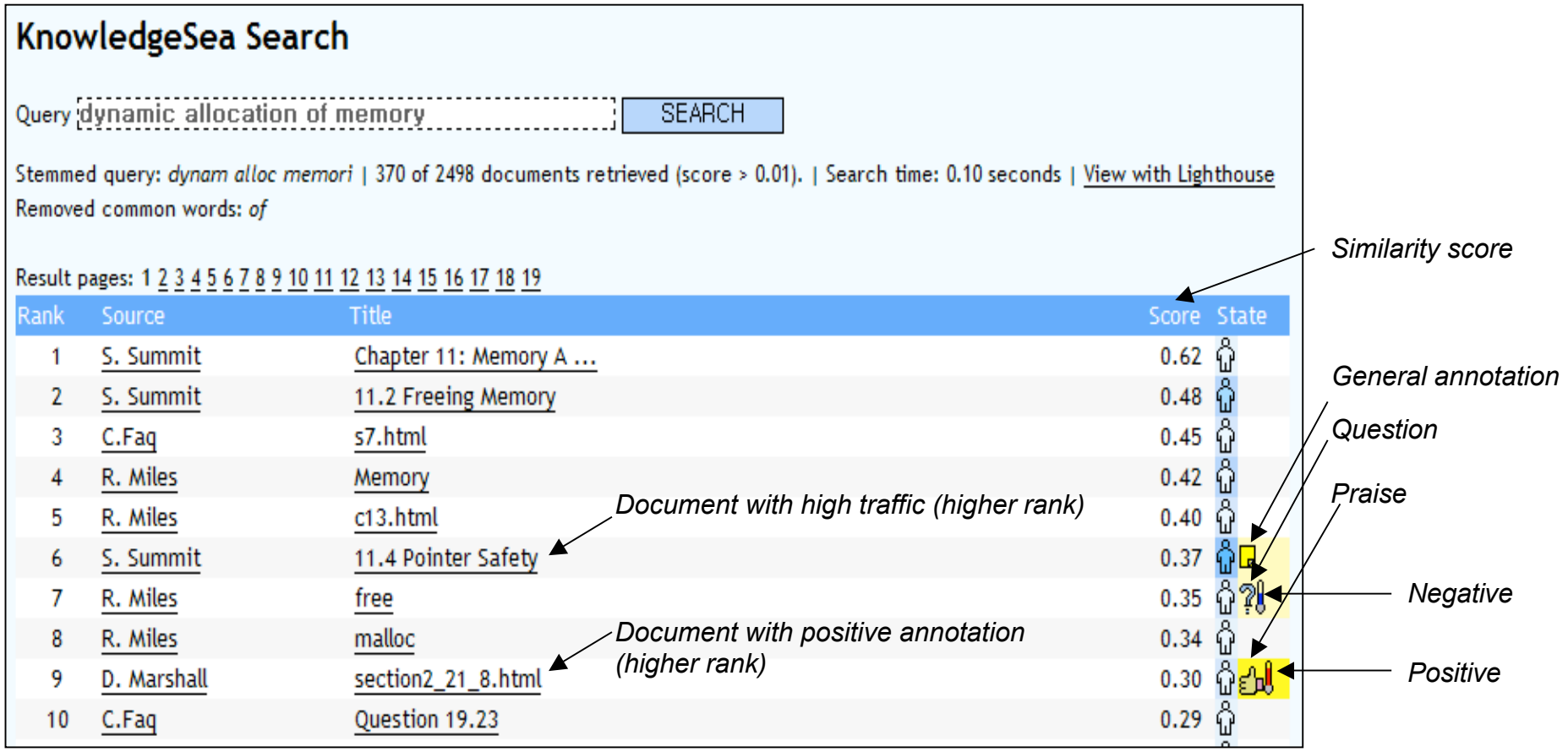


Example of Social Traffic



Example of Social Annotations

System Design and Implementation (cont.)



Research Design

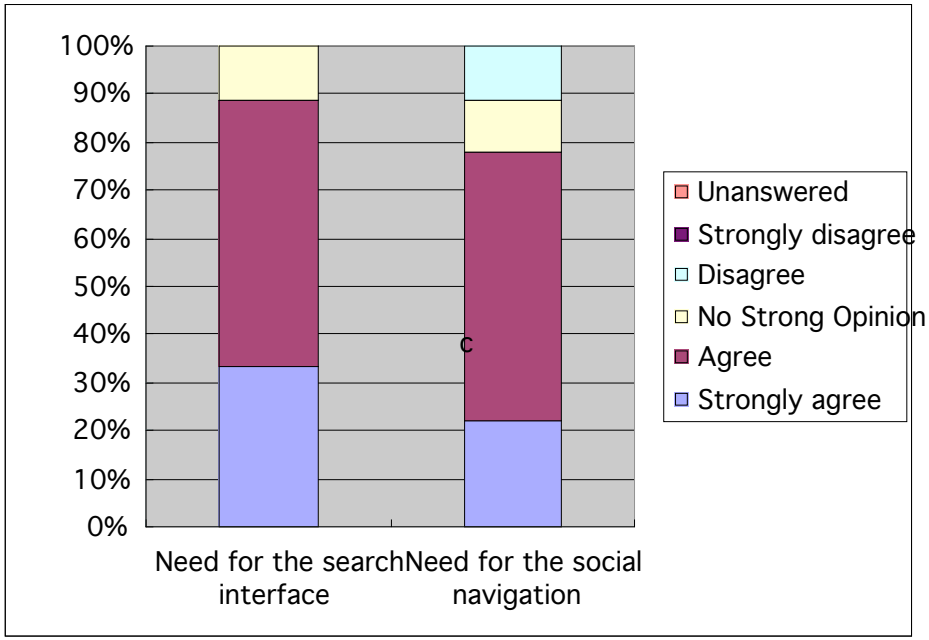
- Hypotheses
  1. Users will **need the social search** capability and will **use it meaningfully**.
  2. Users will **actively select documents with higher social navigation scores**. They ma y select **lower ranked** documents with high group traffic and/or positive annotations.

Research Design - Methodology

- INFSCI 0012 Introduction to Programming course
- Survey
  - Search interface is importa nt?
  - Social navigation support f or search is important?
  - *Hypothesis 1*
- Log analysis
  - 2 months (10/19/04~12/18/04)
    - Number of times search was used
    - *Hypothesis 1*
  - 1 month (11/16/04~12/18/04)
    - Rank, Similarity, Doc I D, Query string, Traffic and Annotation informat ion
    - *Hypothesis 2*

Survey results

- Number of Answers
  - 9 students
- Need for the search
  - **88.9% agreed**
  - 11.1% neutral
- Need for the social navig ation
  - **77.8% agreed**
  - 11.1% neutral
  - 11.1% disagreed
- *Supports Hypothesis 1*



Transaction log analysis results

- Traffic
  - Corresponding to survey results
  - Search service is used
  - Slightly more selection count with social navigation
  - *Supports Hypothesis 1*
- Rank
  - Higher rank with social navigation
  - *No support for Hypothesis 2*

Map	Browsing	**Searching**	Total
299 (36.2%)	423 (51.1%)	**105 (12.7%)**	827

Table 1. Number of times used for each mode

	With group traffic or positive annotations	Without group traffic or positive Annotations
Average rank	**6.48**	8.54
Selection count	**29**	24

Table 2. Average rank and selection count

How attractive are the cues?

- Method of evaluation: beating the random

Cues	Random	Effective
High rank (1-3)	0.15 (3/20)	0.30 (16/53)
Visible traffic	0.08	0.32
High traffic	0.05	0.19

Activity Increase

- Number of documents viewed per query
  - Session length
  - Users viewed more do cuments when they re ceived result sets cont aining group traffic or positive annotations
  - *Supports Hypothesis 1*

	Average
With group traffic	**2.69**
Without group traffic	2

Table 3. Average number of documents viewed per query

	Average
With positive annotations	**4.5**
Without positive annotations	1.94

Table 4. Average number of documents viewed per query

Conclusions

- Implemented and tested the possibility of social search
- *Hypothesis 1*
  - Users agreed with the need for social search
    - Survey results
  - Users in reality used social search services
    - Frequency of usage
- *Hypothesis 2*
  - Social Visual Cues are taken into account
    - Social Navigation is twice as more “attractive” in influencing user na vigation decision than high rank
  - Social visual Cues provide higher prediction for page quality that high rank