

# Facilitating Exploratory Search by Model-Based Navigational Cues

Wai-Tat Fu, Thomas G. Kannampallil, and Ruogu Kang

Applied Cognitive Science Lab

University of Illinois at Urbana-Champaign

405 N. Mathews Ave., Urbana, IL 61801

wfu@illinois.edu

## ABSTRACT

We present an extension of a computational cognitive model of social tagging and exploratory search called the *semantic imitation model*. The model assumes a probabilistic representation of semantics for both internal and external knowledge, and utilizes social tags as navigational cues during exploratory search. We used the model to generate a measure of information scent that controls exploratory search behavior, and simulated the effects of multiple presentations of navigational cues on both simple information retrieval and exploratory search performance based on a previous model called *SNIF-ACT*. We found that search performance can be significantly improved by these model-based presentations of navigational cues for both experts and novices. The result suggested that exploratory search performance depends critically on the match between internal knowledge (domain expertise) and external knowledge structures (folksonomies). Results have significant implications on how social information systems should be designed to facilitate knowledge exchange among users with different background knowledge.

## Author Keywords

Semantic Imitation, SNIF-ACT, exploratory learning, knowledge exchange, social tagging

## ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Collaborative computing. H5.4. Information interfaces and presentation (e.g., HCI): Hypertext/Hypermedia.

## General Terms

Design, Theory, Performance

## INTRODUCTION

In the domain of computer science, representations of data are known to determine implementation of control structures and computations. Similarly, in the domain of

cognitive science, representations and processes of mental contents and skills are shown to have significant impact on human cognitive activities, development, and acquisition of knowledge structures. These observations lead naturally to the thesis that different interface representations and interaction methods will have significant impact on shaping the structure, performance, and emergent behavior of human-computational systems.

The goal of this paper is to show how a computational cognitive model of social tagging and exploratory search can generate useful predictions on the effects of different presentations of navigational cues on facilitating exploratory search and knowledge exchange. Results are useful for informing engineering decisions on how navigational cues should be presented to make knowledge search more effective or efficient, or, in other words, to create more intelligent interfaces.

## Social Tags as Navigational Cues

Recent advances have made the Web a more participatory social-computational systems that allow people to explore, learn, and share information with others. A good example is the increasing popularity of social bookmarking systems such as del.icio.us CiteULike.org, and Bibsonomy.org, which allow users to annotate, organize and share their web-based resources using short textual labels called tags. Many have argued that social tagging systems can provide navigational cues or “way-finders” [6, 10, 14, 15, 20, 24] for other users to explore information. The notion is that, given that social tags are labels that users create to represent topics extracted from Web documents, interpretation of these tags should allow other users to predict contents of different documents efficiently. Social tags are arguably more important in exploratory search, in which the users may engage in iterative cycles of goal refinement and exploration of new information (as opposed to simple fact-retrievals), and interpretation of information contents by others will provide useful cues for people to discover topics that are relevant.

One significant challenge that arises in social tagging systems is the rapid increase in the number and diversity of the tags. As opposed to structured annotation systems, tags provide users an unstructured, open-ended mechanism to annotate and organize web-content. As users are free to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IUT'10, February 7–10, 2010, Hong Kong, China.

Copyright 2010 ACM 978-1-60558-515-4/10/02...\$10.00.

create any tag to describe any resource, it leads to what is referred to as the *vocabulary problem* [9]. Because users may use different words to describe the same document or extract different topics from the same document based on their own background knowledge, the lack of a top-down mediation may lead to an increase in the use of incoherent tags to represent the information resources in the system. In other words, the inherent "unstructuredness" of social tags may hinder their potential as navigational cues for searchers because the diversities of users and motivation may lead to diminishing tag-topic relations as the system grows.

### **Descriptive Models of Social Tagging**

Despite this potential vocabulary problem, recent research has found that at the aggregate level, tagging behavior seemed relatively stable and that the tag choice proportions seemed to be converging rather than diverging [2, 11]. While these observations provided evidence against the proposed vocabulary problem, they also triggered a series of research investigating how and why tag proportions tended to converge over time.

One explanation for the stability was that there was an inherent propensity for users to "imitate" word use of others as they create tags. This propensity may act as a form of *social cohesion* that fosters the coherence of tag-topic relations in the system, and leads to stability in the system. Golder and Huberman showed that the stochastic urn model by Eggenberger and Polya [3] was useful in explaining how simple imitation behavior at the individual level could explain the converging usage patterns of tags. Specifically, convergence of tag choices was simulated by a process in which a colored ball was randomly selected from an urn and was replaced in the urn along with an additional ball of the same color, simulating the probabilistic nature of tag reuse. The simple model, however, does not explain why certain tags would be "imitated" more often than others, and therefore cannot provide a realistic mechanism for tag choices and how social tags could be utilized as navigational cues during exploratory search, not to mention the obviously over-simplified representation of individual users by balls in an urn.

The memory-based Yule-Simon (MBYS) model of Cattuto [2] attempted to explain tag choices by a stochastic process. They found that the temporal order of tag assignment has an impact on users' tag choices. Similar to the stochastic urn model, the MBYS model assumed that at each time step a tag would be randomly sampled: with probability  $p$  the sampled tag was new, and with probability  $1-p$  the sampled tag was copied from existing tags. When copying, the probability of selecting a tag was assumed to decay with time, and this decay function was found to follow a power law distribution. Thus, tags that were recently used had a higher probability of being reused than those used in the past. One major finding by Cattuto et al. was that semantically general tags (e.g., "blog") tended to co-occur more frequently with other tags than semantically narrower tags (e.g., "ajax"), and this difference could be captured by

the decay function of tag reuse in their model. Specifically, they found that a slower decay parameter (when the tag is reused more often) could explain the phenomenon that semantically general tags tended to co-occur with a larger set of tags. In other words, they argued that the "semantic breadth" of a tag could be modeled by a memory decay function, which could lead to different emergent behavioral patterns in a tagging system.

### **A Predictive Process Model of Social Tagging**

Results from previous models were based on analyses of word-word relations as revealed by the various statistical structures in the organization of tags (e.g., how likely one tag would co-occur with other tags or how likely each tag was reused over time). These models are therefore *descriptive* models at the aggregate level, and have little to offer about *predictions* at the level of interface interactions and cognitive processes of individual. To this end, we attempted to adopt a slightly different approach. Rather than describing aggregate behavioral patterns, we drew on research results from the domain of cognitive science to develop a computational cognitive model that characterized knowledge representation and tag choices at the individual level, then studied how their effects could generate useful predictions on search performance.

In our previous work [6-8], we argued that, rather than imitating other users at the word level, one possible explanation for this kind of social cohesion could be grounded on the natural tendency for people to process tags at the semantic level, and it was at this level of processing that most *imitation* occurred. This explanation was supported by research in the area of reading comprehension [17], which showed that people tended to be influenced by meanings of words, rather than the words themselves during comprehension. Assuming that background knowledge of people in the same culture tend to have shared structures (e.g., using similar vocabularies and their corresponding meanings in order to conform and communicate with each), users of the same social tagging system may also share similar semantic representations of words and concepts, even when the use of tags may vary across individuals at the word level. In other words, we argued that part of the reason for the stability of social tagging systems can be attributed to the shared semantic representations among the users, such that users may have relatively stable and coherent interpretation of information contents and tags as they interact with the system. Based on this assumption, we developed the *semantic imitation model* [7, 8] that predicts how different semantic representations may lead to differences in individual tag choices and eventually different emergent properties at the aggregate behavioral level. The model also predicts that the *folksonomies* (i.e., knowledge structures) in the system reflect the shared semantic representations of the users.

In the current paper, we propose an extension of the semantic imitation model to predict how different presentations of navigational cues may influence search

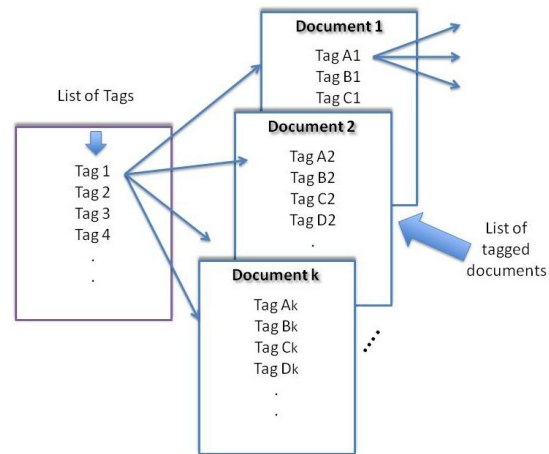
performance. We will first describe performance in an exploratory search task and contrast it with a standard information retrieval task using traditional search engines. We will then introduce the details of the semantic imitation model and how it can be applied to characterize social tagging and search behavior. We will then describe how the model can be combined with our previous SNIF-ACT model of Web search [5] to generate predictions on the effects of different presentations of navigational cues on search performance. We will also show how this potential improvement in performance may differ in systems that are mostly used by domain experts and novices.

### SIMULATING SEARCH PERFORMANCE

Figure 1 shows a scenario in which a user is performing a search task in a simplified social tagging system. For simple information retrieval, the searcher may utilize some keywords to locate a single piece of information (e.g., address of a movie theatre). For exploratory search, the searcher may be looking for information related to a particular topic (e.g., financial crisis), but he or she may not know exactly where and what to look for information relevant to this topic. The searcher can begin by inputting the topic and selecting one of the tags that are associated with the topic as suggested by the social tagging system. The searcher may then click on one of the tags to access the list of documents associated with the tag. Each of the documents will also contain a list of tags created by other users, so the searcher can use these tags to evaluate whether the document may contain information relevant to their information goal, as well as to gain some ideas about what other tags/keywords may be related to the topic that he or she is searching for. In contrast to information retrieval, in exploratory search the information that the searcher is looking for is seldom contained in a single document. Rather, the searcher will likely need to collect and integrate a set of documents related to the topics (and subtopics) during their search, and may need to iteratively refine their search goals to learn more about the topic as they navigate in the system [6-8, 15, 16, 20].

In many ways, the tag-based exploratory search as depicted in Figure 1 is similar to information retrieval in a traditional search engine, in which a user types in a keyword and a list of documents will be returned from the search engine. There are, however, three main differences between searching in a traditional search engine and in a social tagging system. First, in a social tagging system, users can browse through tags created by others to pick one that is closest to the topic that they are interested in, but in a traditional search engine users have to come up with their own keywords as they browse through the documents. Second, the list of documents returned from a social tagging system are determined by tags created by other users, but in a traditional search engine the links between keywords and document are usually calculated by machine-learning algorithms, which may or may not be capturing the effect of social cohesion in a tagging system as discussed earlier.

Third, tags associated with the list of returned documents can help users to refine their information goal by interpreting the association of these tags (and the related topics) within and across documents, such that they can gain a better understanding of *not only* the topics that they are looking for, but *also the relations among these related topics* as they are distributed across documents tagged by users of the system [16]. In contrast, in traditional search engines, the list of documents returned are usually not semantically related and thus they do not inherently support this kind of *sensemaking* activity [6, 8, 10, 14, 16, 18, 20].



**Figure 1. An example of exploratory search in a social tagging system. The list of tags are provided by the system, when a user clicks on a tag, a list of documents associated with that tag will be presented. Each of these documents also has a list of tags associated with it.**

When the searcher is retrieving a single piece of information, search performance can be measured by the number of documents browsed before the target document can be found (the fewer the number of documents browsed, the better the performance). However, because in exploratory search the goal is not to find a single document, this measure of performance cannot be directly used. Instead, one can assume that the number of documents that the searcher collects (or saves as bookmarks) in a certain amount of time can be used as a measure of exploratory search performance. In other words, if system A allows a searcher to find more relevant documents in the same amount of time (or clicks) than system B, we assume that system A provides better support for exploratory search than system B. We will use this measure for exploratory search performance in our simulation.

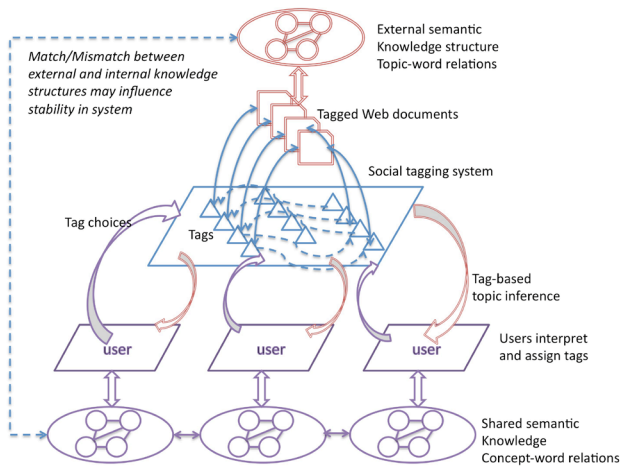
### Relation to Previous Information Search Models

Before discussing our model of tag-based exploratory search, it will be useful to review briefly previous models of information search [5, 22-24]. One important control variable in these models is the measure of information scent [5, 22]. Information scent is a general term describing the perception of the relevance of a link to the information goal of the searcher based on information cues presented by the system (e.g., link texts on the search-result page returned

from either a search engine or social tagging system). There are different calculations of the measure of information scent, but most of them are derived from some forms of statistical text processing techniques that calculates the semantic distance between two words. Results from previous models suggest that the use of information scent as a control variable for searching behavior provides good match to navigational behavior.

### THE SEMANTIC IMITATION MODEL

In this section, we will briefly review the major components of the semantic imitation model [7, 8] that explains the process of social tagging based on semantic interpretation of content. The model has properties that allow us to separately model the external and internal knowledge structures and effectively integrate it with a psychologically plausible choice mechanism. The model assumes that when users navigate in a social tagging system, they *look at existing tags* (created by previous users), *infer topics* related to the existing tags based on the semantic interpretation of the tags, *select the tags* that are most relevant to the topics that the user is interested in, *comprehend the content* of the document (resource) that is being tagged, and then *choose a tag* (or reuse an existing) that is appropriate for the resource (see Figure 2). We will first focus how the model represent semantic knowledge. We will then describe how the model interpret and create tags based on the semantic knowledge representations, before we describe how the model can simulate exploratory search in a social tagging system.



**Figure 2. Semantic Imitation Model: Users interpret tags based on their knowledge and infer topics in a document and choose tags to represent the latent semantics in the documents.**

### Representation of semantic knowledge

There are three main components in the model: words, semantic concepts (or topics) underlying the words, and documents that contain both the words and the topics. It is assumed that concepts can be represented as a probability distribution of words. For example, the concept “health” can instantiate many associated words such as hospital, doctor, surgery, etc. If  $c$  represents the set of available concepts and  $w$  represents the set of words, then  $p(w|c)$  is

the probability distribution of words, given a set of concepts. Using Bayes theorem, we can then calculate the set of concepts given a set of words.

$$p(c_k | w) = \frac{p(\bar{w} | c_k) p(c_k)}{\sum_i p(w | c_i) p(c_i)} \quad (1)$$

We assume that for each concept,  $c$ , the associated word in the multinomial distribution  $p(w|c)$  is normally distributed. Thus, there will be words in a concept that are more “central” than others. For example, for the concept “money”, the word “dollar” is more central than the word “exchange”. By assuming that concepts have overlapping normal distributions over words, we can simulate different degrees of ambiguities of the semantic meanings of words.

One major advantage of the current semantic representation is that it is a generative model, in the sense that we can generate documents based on the assumed set of word-topic distributions. This means that we can use a unified framework to simulate both the internal and external knowledge structures in the system. For example, given a particular distribution of topics and words in a set of documents, we can simulate the differences in internal knowledge structures between experts and novices by varying the match between internal and external knowledge structures (i.e., experts will have a better match to the word-topic distributions in the documents than novices). Previous research has shown that the model can produce documents that match well with real documents in large corpora [13].

In the current simulations, we manipulated the standard deviations of the normal distributions of the word-concept distributions of the users. When the standard deviation was set to a higher value, there would be a decrease in the specificity of word use to represent a concept (e.g., when creating tags) as well as in word interpretation (e.g., when comprehending the meaning of words/tags). In other words, we assumed that experts will have higher specificity in both their word use and interpretation than novices. Note that this definition is consistent with research that adopts a network approach to define expertise [21].

### Interpreting tags

As the user navigates through the system, existing tags of a resource act as retrieval cues for inferring the related concepts. As users browse through a set of tags that were previously created for a resource, semantic representations for those tags (words) are activated. This tag-based topic inference can be represented as a measure that estimates the probability that a set of tags will instantiate a specific set of topics (or concepts),  $p(c|w)$ , where  $c$  is the set of concepts and  $t$  are the tags that are associated with a resource. Substituting tags,  $t$  into equation (1) we get,

$$p(c_k | t) = \frac{p(t | c_k) p(c_k)}{\sum_i p(t | c_i) p(c_i)} \quad (1a)$$

Equation 1a represents the probability that a given tag can predict a particular concept. This tag-based topic inference process will facilitate the evaluation of the relevance of the document, as well as the comprehension of the document if a document is selected, as we will discuss next.

### Comprehension of document contents

The comprehension process consists of combining the concepts that the user abstracts from the tags and then combines it with the concepts extracted from the document. This final set of topics that are obtained uses the prior distributions of all concepts represented by  $p(c_k|t)$  (calculated from 1a). This is represented by the posterior probability  $p(c_k|d)$  extracted from document,  $d$ .

$$p(c_k | d) = \frac{p(\bar{w} | c_k) p(c_k | t)}{\sum_i p(\bar{w} | c_i) p(c_i | t)} \quad (2)$$

### Assigning Tags

Tag creation is based on concept-word and word-concept relations. In order to do this, the probability of selecting a new tag given the set of existing tags and the words in the document is computed. In (3),  $w_{new}$  is the new tag created by the user and  $d$  represents the aggregate of all words in the document and the existing tags.

$$p(w_{new} | d) = \sum_i p(w_{new} | c_k) p(c_k | d) \quad (3)$$

Thus,  $p(w_{new}|d)$  is the likelihood that the words and tags associated with a document will predict the creation of a new tag,  $w_{new}$ . An important element of the tag choice process is that it should mirror a behavioral decision making process. For this we used a random utility model (RUM) that has been used extensively to model human choice behavior [19]. We use the degree of representativeness of a tag as the measure of utility of the tag. The degree of representativeness of a tag provides a measure of how much a tag represents the semantic contents of a document. We also use a random variable,  $\sigma$  to incorporate a degree of uncertainty into the model.

$$U_w = p(w | d) + \sigma \quad (4)$$

where  $\sigma$  follows a double exponential distribution, and it can be shown that the probability that  $U_w$  is the maximum can be expressed as ( $b$  controls how likely a word with low utility will be chosen over one that has high utility)

$$p(U_w > U_j \text{ for all } j) = \frac{\exp(U_w/2b)}{\sum_j \exp(U_j/2b)} \quad (5)$$

A new tag assignment occurs only when the tag that has a higher utility was more representative than the existing tags.

$$\frac{\max[U_w]}{\max[U_i]} > h \quad (6)$$

In (6),  $\max[U_w]$  is the maximum utility among all words and  $\max[U_i]$  is the maximum utility among all existing tags. The assumption here is that users create new tags reflect the content of a document. Though in this paper we focus only on the informational value of the tags, previous research has shown that representativeness reflects the association between cues (a tag in this case) and an item retrieved from personal memory. Thus, it is possible to take into account cases where people use personal tags (e.g., to read) to qualify a resource.

As opposed to other models of social tagging, the semantic imitation model was based on a cognitively plausible tag choice mechanism that was coupled to the formal representations of semantic knowledge that exist in both external documents and internal knowledge structures of the users. The model provides an explanation on generally accepted concepts of tag convergence and stabilization based on a cognitive model of individual taggers (users). Additionally, it has a mechanism for generating testable predictions about behavioral patterns in systems generated by different user populations (e.g., experts vs. novices).

### Exploratory search

To model exploratory search, we simulated a searcher who is searching for information related to a randomly selected topic. Because one can only measure the actual word use by the searcher, we cannot use (1a) to directly calculate information scent. Instead, when the searcher enters the first tag to search for or select a tag during navigation, we can use that to infer what this person may be looking for by summing over all possible topics that are related to the selected tag (see 3). Therefore, the measure of information scent of each of the tag  $IS(tag)$  can be calculated by

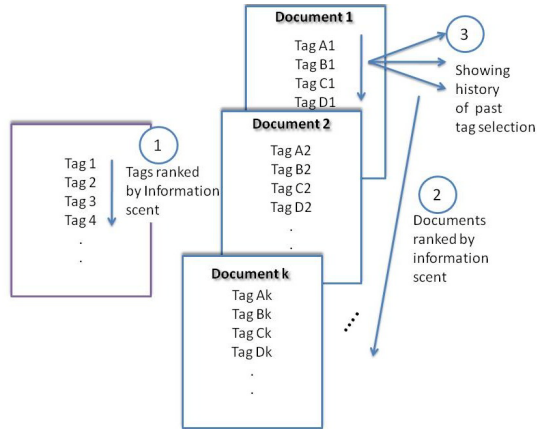
$$\begin{aligned} IS(tag) &= p(tag|tag_{selected}) \\ &= \sum_k p(tag|c_k) p(c_k|tag_{selected}) \end{aligned} \quad (7)$$

in which  $tag_{selected}$  represents the last tag selected, and  $tag$  represents each of the tag to be presented. We used  $IS(tag)$  as a measure of how the searcher will perceive that a tag is relevant during exploratory search. However, in our previous work as well as others, it was found that the sequential order of links (or in this case, tags) had significant effect on search performance [5]. Specifically, it was found that people tended to engage in local cost-benefit tradeoffs [4] as they evaluated links sequentially, such that they would click on a link positioned at the top even when there was a higher-scent link positioned at a lower position.

### Model-based Presentations of Navigational Cues

Figure 3 shows the three model-based presentations of navigational cues in a social tagging system. First, we simulated exploratory search performance in a system that presents tags based on its information scent. Under the assumption that tags are created to represent topics in documents, ranking tags based on their information scent should also increase the efficiency of search (in terms of amount of relevant information found per unit time).

In the simulation, we assumed that the model-searcher would start with a keyword (for information retrieval) or a word sampled from a topic (for exploratory search). This word would be used to generate a list of related tags by the system. In the *random arrangement* condition, the order of these tags will be randomized. Otherwise, the tags would be presented in the order of their information scent.



**Figure 3. The three model-based presentations to be tested by the mode simulation during the exploratory search process: (1) ranking tags based on their information scent, (2) ranking documents based on the average information scent of tags, and (3) leaving a document when the information scent drops below threshold (i.e., satisfying).**

Following the SNIF-ACT model [5], selection of tags would depend on *both* the sequential order of the tags *and* its information scent. Specifically, there were two competing actions as the model evaluated each tag and decide which tag to click. The *evaluate-next* action would move on to the next sequential link, and the *select-tag* action would click on a tag. The utilities for each action were calculated based on the Bayesian satisficing model (BSM, see [4, 5]) as:

$$U_{ev}(n+1) = \frac{U_{ev}(n) + IS(tag)}{1 + N(n)}, U_s(n+1) = \frac{U_s(n) + IS(best\ tag)}{1 + k + N(n)} \quad (8)$$

where  $U_{ev}$  and  $U_s$  were the utilities for the action evaluate-next and select-tag respectively.  $n$  represents the model cycle. After each cycle, the utilities would be updated according to (8).  $IS(best\ tag)$  represented the information scent of the tag that had the highest IS value among tags that were evaluated so far.  $N(n)$  was the number of tags evaluated up to cycle  $n$ .  $k$  is a constant that controls the initial bias for the model to evaluate more links (i.e., initially  $U_s$  will be lower than  $U_{ev}$ , but as  $N(n)$  increased, this difference diminished, such that the model would be biased to evaluate at least a few tags before selecting a tag). These two actions competed with each other, and at any model cycle one action would be selected based on the utilities of the actions (see the softmax equation in eqn 5).

When a tag was selected, the model would read through documents that were associated with the selected tags. When the model-searcher read a document, it would judge whether the document contained a relevant topic based on the calculation of  $p(c_k|d)$ . Specifically, the document would be saved if

$$p(c_k|d) > H_{relevant} \quad (9)$$

We set  $H_{relevant} = 0.5$  throughout the simulation. Once a document was read (regardless it was saved or not), the model would go back to the previous page and continued with the next tag, and so on.

In the second model-based presentation, in addition to ranking tags according to their information scent, we also simulated the situation in which the interface also ranked the presentation of documents returned from a tag selection. The ranking was, however, based on the weighted sum of information scent between the selected tag and the set of tags associated with each of the documents. Following the SNIF-ACT model, the information scent of a document was calculated by

$$IS(document) = \sum_j IS(tag_j) * \exp(-dj) \quad (10)$$

in which  $d$  represents a decay parameter and was set to 0.5 as in [17], and the summation was for all tags  $j$  associated with the document. The exponential weighting was to control for the additive effect of the number of tags in a document to the measure of  $IS(document)$ .

Similar to the evaluation of tags, we also assumed that both the sequential order of documents and their information scent would influence how likely the documents would be read. The document selection process was similar to that in tag selection, and the same utility equations in (8) were used, except that  $IS(tag)$  was replaced by  $IS(document)$ .

The final model-based cue was to show the history of tag selection as the model navigated in the system. We assumed that when the history of tag selection was included, the searcher could have more integration of topics browsed during the navigation, and would be able to better distinguish what information was relevant. In addition, we assumed that the calculation of information scent would more accurately reflect the relevance of tags and documents when the history of tag selection was taken into account. Specifically, when subsequent tags were selected, we calculated information scent of each tag by including all tags selected in (7). Tags and documents would then be ranked according to the extent to which they matched the whole history of tags selected by the model-searcher.

## SIMULATING THE MODEL

In this section, we describe the simulation set up for the semantic imitation model for testing the overall properties of folksonomies created by the semantic imitation model by experts and novices, as well as the exploratory search performance of an expert and novice model-searcher. We will first describe the basic simulation set up below.

### Basic Simulation Set Up

Following the method by others [1, 13], a set of 100 documents was generated. Each document had a random set of topics assigned to it. The topics were randomly sampled from a set of 100 sample topics. As explained earlier, for each topic a set of words were sampled from a multinomial distribution of 2500 words. Each word had a prior probability that was normally distributed with a standard deviation of 1. Each document had 500 words.

At the start of simulation, a document was randomly selected. Using an unbiased topic inference process the probabilities for each of the 100 topics were computed. The probabilities for each of the topics were calculated based on the available words ( $w$ ) in the document. This was the  $p(c_k|w)$  for the document (for topics  $k=1$  to 100)(see equation (1)). It was also assumed that at the start of the simulation there was no tag assigned to any document. The set of  $p(c_k|w)$  for all documents was used to compute the  $p(w_i|w)$  for all words in the vocabulary. The utilities of all possible words in the vocabulary were then calculated (using equation (4)).

In the second stage of the simulation, the tag-based topic inference process was invoked by the existing tag, which semantically primed the topic extraction and comprehension process. In order to achieve this, each  $p(c_k)$  would be substituted by  $p(c_k|w)$  from the previous iteration of equation (1). Additionally, instead of  $w$ , the previously assigned tag  $t$  would be used. The probability values of  $p(c_k|t)$  would then be updated for all topics and used as the prior distribution of concepts during comprehension (using equation (2)). Utilities of all words would be calculated and the word that has the maximum utility was compared to the maximum utility of the existing tags. If the ratio exceeded the threshold parameter,  $h$ , the new word was added as a tag to the document. This process continued across 100 iterations for all documents. The threshold parameter  $h$  was set at 1 and  $b$  was set to 0.01 in all simulation runs.

### The Expert and Novice Network

The major benefit of the semantic imitation model was its flexibility in creating different knowledge representations for users. As described earlier, by changing the spread of prior distributions of words over all the available words, different knowledge representations of the user could be created. The smaller spread (i.e., lower s.d.) in the probability distribution of words within each topic implied that the words were more accurate in predicting the concepts in the document, such that the simulated user would be better able to interpret a tag and infer the topic as well as to assign a tag to represent the topic. We assumed that this reflected the performance of domain experts.

The purpose of this simulation was to determine the differences in the overall properties of tags (i.e., the resulting folksonomy) created by experts and novices. The basic simulation parameters were the same as the basic set up (100 documents, 100 topics, 2500 words). There were two conditions that were tested in this simulation. In the

*expert-network*, 35 experts were simulated to create tags in the system, while in the *novice-network*, 35 novices were simulated to create tags in the system. The following properties were computed at the end of each simulation run: total number of tag applications (a tag application was the addition of a tag to a document), number of tags that were reused, number of unique tags that were created, and the variance in the number of tag applications.

### The Expert and Novice Searcher

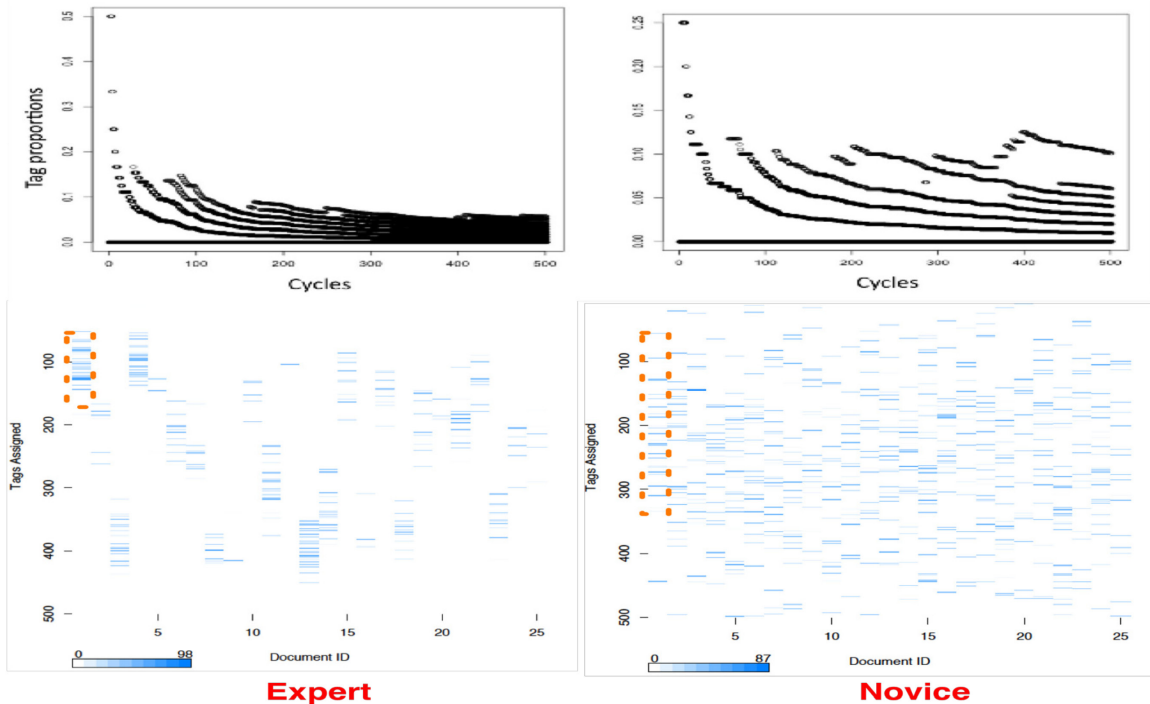
We also simulated search efficiency in both networks with different model-based presentations of navigational cues shown in Figure 3. We created the expert and novice searcher, each of which performed a simple document retrieval task and an exploratory search task in each network. We therefore had four different sets of simulation results from the combinations of expert and novice models searching in the expert and novice networks. In each set, separate simulations were conducted for each addition of the model-based presentations (i.e., they were cumulative). In addition, a condition in which tags and documents were randomly organized was simulated as a control condition. Therefore, we simulated performance in each of the four conditions in each of the four sets of simulations.

## RESULTS

### General Properties: Convergence, Stabilization of Tags

The most generally accepted property of social tagging systems is that the proportion of tags assigned to a document converges over time [12]. So, as the total number of tags increase in a system, the ratio of the frequency of a tag to the total number of tags remains fairly constant. This emergent property of tags, called *convergence*, was attributed to the social nature of the tagging process. In our previous simulations [7], we showed that the semantic imitation model produced not only the convergence, but also predicted how experts and novices could lead to different rates of convergence. To provide the background for the effects of the model-based presentation, we showed the results here to highlight the differences.

The two graphs on the top half of Figure 4 shows simulation results generated from the model. The graphs are based on simulation of 100 documents with each point representing the proportion of a tag (y-axis) at a certain time point during the simulation (x-axis). Similar to results obtained by Golder and Huberman [9], the tag proportions created from the semantic imitation model converged over time. The convergence of tags in the semantic imitation model can be explained as follows: users tag choices are driven by the degree of representativeness of a tag to the concepts extracted from the document. The extraction of concepts is influenced by the semantic interpretation of existing tags (rather than a direct imitation). The commonality in the semantic representation of words and concepts among all users will lead to a coherent interpretation and choice of tags that are perceived to be most representative of the documents.



**Figure 4. Convergence and stabilization in tags created by experts (top-left) and novices (top-right). The corresponding spread of tags created by experts and novices are shown at the bottom panels.**

The faster convergence in the expert network can be explained as follows: the tags assigned by experts were more predictive of the topics in the document and experts could extract these topics better than novices. Additionally, other experts tagging the same resource tended to choose the same higher quality tags. In contrast, novices were less knowledgeable about the contents of the document and consequently less effective in extracting the appropriate topics (and therefore tags) from the documents. Novices therefore selected tags that were more diverse than experts and hence the slower convergence.

This phenomenon was further clarified in the graphs shown on bottom half of figure 2. The x-axis represents a subset of documents (25) from the 100 documents used for the simulation and the y-axis shows a subset of tags (500) from the total of 2500. The “spread” of the tags created by experts was much narrower than those created by novices (shown within the dashed box for one document). This means that experts’ tag creation process was similar in nature and they reached a consensus much quicker than novices. In contrast, tags created by novices were more diverse, leading to slower convergence.

#### Performance on Simple Information Retrieval

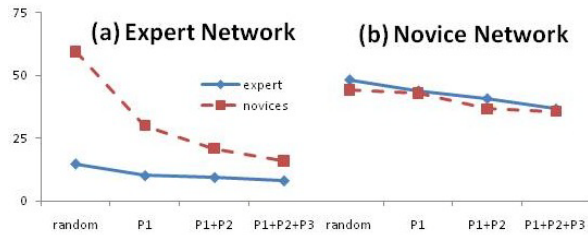
We simulated the model's performance on information search for a single document to understand how different presentations of results could benefit experts and novices in the expert and novice network. Each model searcher would be given a keyword to search for a specific document in the network. The system would then return a set of tags most related to the keyword. The model would then select the tag

sequentially based on the order presented by the system (which varied depending on the condition). We then tabulated the number of documents browsed before the document was found for each condition, and average the results over 100 iterations.

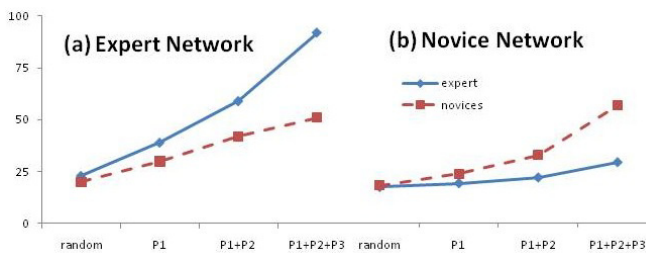
Figure 5 shows the results for the expert and novice searcher in the networks. In the expert network, the expert searcher was much better at finding the document than novices. The three model-based navigational cues seemed to have helped the novices significantly in finding the document, as the number of documents browsed was much reduced when ranked tags were presented, and this number decreased even more as presentations of documents were ranked, and when tag history was presented to rank the documents. Given that the performance of the experts were already good even when tags were randomly presented, the improvement was smaller than that for the novice. However, we did see improvement as more model-based navigational cues were used.

The results showed that in the expert network, when the “quality” of tags was good, experts were very good at utilizing the tags as navigational cues to find the target document. Indeed, experts could find the target document much more efficiently than novices even in the random arrangement of tags. This suggested that experts were much better at evaluating which tags were good than novices (such that there was a much higher difference in the  $IS(tag)$  values for experts than novices, making them better at selecting the right tag to search as specified in eqn 8). On





**Figure 5. In simple information retrieval, mean number of documents browsed before the desired document was found in each of the presentation conditions for expert and novice searchers in the (a) expert and (b) novice network.**



**Figure 6. In exploratory search, the mean number of documents saved in each of the presentation conditions for expert and novice searchers in the (a) expert and (b) novice network.**

the other hand, the model-based navigational cues did help the novices to find the target document more efficiently, suggesting that when novices could not evaluate which tags were good, ranking the high-quality tags for them would significantly help them to navigate to the target document. Indeed, when all three model-based navigational cues were used, novices could perform almost just as well as experts, suggesting that the model-based presentation of expert tags were effective in guiding novices to navigate in the system.

As Figure 5b shows, results were quite different in the novice network. Although the model-based navigational cues did help both expert and novice searchers, the improvement in efficiency was much smaller than that in the expert network. In addition, there was virtually no difference between performance by experts and novices. This suggested that when the quality of tags were low, users could not benefit much from the tags to help them navigate to the target document. In other words, the effectiveness of the model-based presentation techniques were impeded by the inherently low-quality tags generated by novices. Apparently, the ranking of tags and documents based on the noisy estimates of information scent was not very useful in guiding searchers to navigate in the system.

### Performance on Exploratory Search

In exploratory search, the main performance measure was the number of relevant document saved within 100 tag selections. We assumed that the more number of document saved, the better was the performance. Figure 6 shows the number of documents saved during exploratory search for expert and novice searchers in the networks. In the expert network, model-based navigational cues apparently helped both experts and novices search for relevant information. Interestingly, for experts, as more model-based navigational

cues were introduced, the improvement was significantly more than for novices (as shown by the exponential increase in effectiveness in Figure 6a). This suggested that when experts were exploring for information with good quality tags, the model-based navigational cues significantly augmented the exploration of information much more than novices. This could be attributed to the fact that experts were much better at interpreting the tags. Thus, the estimation of information scent for experts was better than novices. Experts could therefore select more relevant tags within the same number of tag search compared to novices. Similarly, the ranking of documents and the inclusion of search history had significantly improved exploratory search performance for the expert searcher, as experts were better at interpreting the tags to navigate to the right documents. This was most salient when the history of tags was used, in which the number of documents found increased almost 4 times as much compared to the random arrangement condition.

Similar to performance in simple information retrieval, exploratory search performance in the novice network was much poorer than in the expert network. Interestingly, in contrast to performance in information retrieval, novices benefited more from the model-based navigational cues than experts. Apparently, novices found more relevant documents than experts when searching in a novice network. This could be attributed to the fact that novices were able to find more documents that had information scent values higher than the threshold than experts, but this was mostly caused by the fact that their estimation of information value was less accurate. Another way to interpret the results was that novices had a better match between their internal knowledge and the external knowledge in the novice network, and thus they were able to "better utilize" the navigational cues. However, given that the quality of tags in the novice network was poorer, the higher number of saved documents could be an artifact of poorer evaluation of relevance by the novices. In other words, novices might be "misguided" by the poorer tags. Nevertheless, both experts and novices did find more relevant information with the model-based navigational cues, suggesting that the presentations were useful in enhancing the navigational values of tags in the system.

### CONCLUSION AND DISCUSSION

We presented an extension of the semantic imitation model to simulate an expert and a novice network, and simulated search performance of an expert and a novice searcher navigating in each of the networks. In general, we found that the model-based navigational cues were useful in facilitating both simple information retrieval and exploratory search, but this effect was more prominent in the expert network than in the novice network. The result suggested that both the quality of tags and the presentation formats of tags and documents could facilitate effective and efficient navigation in a social tagging system. We, however, did not imply that social tags are created only for

navigational purposes, indeed others have showed that there are a number of other possibilities (e.g., [25]). Future research will include these other possibilities.

Our results showed that even when quality of tag was low, simple manipulation in the presentation of information could serve as good navigational cues to guide users to find the right information. Our results also highlighted the value of a predictive process model in generating testable interface features and representations that could significantly augment performance in a system.

The semantic imitation model presented in this paper shows a good example of how theories of cognitive science and artificial intelligence may provide researchers a strong theoretical basis for empirical investigations of behavior in social information systems, and how they can lead to potential design insights for future social tagging systems. The results from the simulations can also lead to testable predictions and guide the design of empirical studies to test the effects of different interface representations and interaction methods on performance.

## REFERENCES

1. Blei, D., A. Ng, and M. Jordan, *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003. 3: p. 993-1022.
2. Cattuto, C., V. Loreto, and L. Pietronero, *Semiotic dynamics and collaborative tagging*. Proceedings of National Academy of Sciences, 2007. 104: p. 1461-1464.
3. Eggenberger, F. and G. Polya, *Über Die Statistik Verketter Vorgage*. Zeit. Angew. Math. Mech., 1923. 1: p. 279-289.
4. Fu, W.-T. and W. Gray, *Suboptimal tradeoffs in information seeking*. Cognitive Psychology, 2006. 52(3): p. 195-242.
5. Fu, W.-T. and P. Pirolli, *SNIF-ACT: A cognitive model of user navigation on the World Wide Web*. Human-Computer Interaction, 2007. 22: p. 355-412.
6. Fu, W.-T., *The microstructures of social tagging: A rational model*, in *Proceedings of the ACM 2008 conference on Computer supported cooperative work*. 2008: San Diego, CA. p. 229-238.
7. Fu, W.-T., T.G. Kannampallil, and R. Kang, *A Semantic Imitation Model of Social Tag Choices*. , in *Proceedings of the IEEE conference on Social Computing*. 2009: Vancouver, BC. p. 66-72.
8. Fu, W.-T., et al., *Semantic Imitation in Social Tagging*. ACM Transactions on Computer-Human Interaction, in press.
9. Furnas, G.W., et al., *The vocabulary problem in human-system communication*. Communications of the ACM, 1987. 30(11): p. 964-971.
10. Furnas, G.W., et al. *Why do tagging systems work?* in *CHI '06 Extended Abstracts on Human Factors in Computing Systems*. 2006. Montréal, Québec, Canada.
11. Golder, S. and B. Huberman, *Usage patterns of collaborative tagging systems*. J. Inf. Sci., 2006. 32(2): p. 198-208.
12. Golder, S.A. and B.A. Huberman, *Usage patterns of collaborative tagging systems*. J. Inf. Sci., 2006. 32(2): p. 198-208.
13. Griffiths, T.L., et al., *Topics in Semantic Representation*. Psychological Review, 2007. 114(2): p. 211-244.
14. Kammerer, Y., et al. *Signpost from the masses: learning effects in an exploratory social tag search browser*. in *Proceedings of the 27th international Conference on Human Factors in Computing Systems CHI '09*. 2009. Boston, MA, USA.
15. Kang, R., et al., *Conformity out of Diversity: Dynamics of Information Needs and Social Influence of Tags in Exploratory Information Search*, in *Proceedings of the 13th International conference of Human-Computer Interaction*. 2009: San Diego, CA.
16. Kang, R. and W.-T. Fu, *Exploratory Information Search by Domain Experts and Novices*, in *Proceedings of the 2010 Conference on Intelligent User Interfaces*. 2010: Hong Kong, China.
17. Kintsch, W., *The role of knowledge in discourse comprehension: A construction-integration model*. Psychological Review, 1988. 95: p. 163-182.
18. Marchionini, G., *Exploratory search: from finding to understanding*. Commun. ACM, 2006. 49(4): p. 41-46.
19. McFadden, D., *Conditional Logit Analysis of Qualitative Choice Behavior*, in *Frontiers of Econometrics*, P. Zarembka, Editor. 1974, Academic Press: New York. p. 105-142.
20. Millen, D., et al., *Social bookmarking and exploratory search*, in *ECSCW 2007*. 2007. p. 21-40.
21. Noll, M.G., et al., *Telling experts from spammers: Expertise ranking in folksonomies*, in *Proceedings of the ACM conference on Information Retrieval (SIGIR)*. 2009: MA: Boston. p. 612-619.
22. Pirolli, P. and S. Card, *Information Foraging*. Psychological Review, 1999. 106(4): p. 643-675.
23. Pirolli, P. *An elementary social information foraging model*. in *Proceedings of the 27th international Conference on Human Factors in Computing Systems CHI '09*. 2009. Boston, MA, USA.
24. Qu, Y. and G. Furnas, *Model-driven formative evaluation of exploratory search: A study under a sensemaking framework*. Inf. Process. Manage., 2008. 44(2): p. 534-555.
25. Sen, S., et al., *tagging, communities, vocabulary, evolution*, in *Proceedings of the 2006 20th anniversary conference on Computer Supported Cooperative Work*. 2006, ACM: Banff, Alberta, Canada.