Augmented Social Cognition: Using Social Web technology to enhance the ability of groups to remember, think, and reason

Ed H. Chi
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304, USA
+1-650-812-4312
chi@acm.org

ABSTRACT

We are experiencing a new Social Web, where people share, communicate, commiserate, and conflict with each other. As evidenced by systems like Wikipedia, twitter, and delicious.com, these environments are turning people into social information foragers and sharers. Groups interact to resolve conflicts and jointly make sense of topic areas from "Obama vs. Clinton" to "Islam."

PARC's Augmented Social Cognition researchers -- who come from cognitive psychology, computer science, HCI, CSCW, and other disciplines -- focus on understanding how to "enhance a group of people's ability to remember, think, and reason". Through Social Web systems like social bookmarking sites, blogs, Wikis, and more, we can finally study, in detail, these types of enhancements on a very large scale.

Here we summarize recent work and early findings such as: (1) how conflict and coordination have played out in Wikipedia, and how social transparency might affect reader trust; (2) how decreasing interaction costs might change participation in social tagging systems; and (3) how computation can help organize user-generated content and metadata.

Categories and Subject Descriptors

H.5.3 [Information Interfaces]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work, Web-based interaction; H.3.5 [Information Storage and Retrieval]: Online Information Services; H5.2 [Information interfaces and presentation]: User Interfaces; K.4.3 [Computers and Society]: Organizational Impacts – Computer-supported collaborative work; H3.3 [Information Search and Retrieval]: Relevance Feedback, Search Process, Selection Process.

General Terms

Measurement, Performance, Design, Economics, Experimentation, Human Factors.

Keywords

Social Web, Augmented Social Cognition, social system, CSCW, HCI, research methods, Wikipedia, delicious, social tagging, characterization, modeling, summary, overview.

SIGMOD'09, June 29–July 2, 2009, Providence, RI, USA.

1. INTRODUCTION

One enduring core value of research in Human-Computer Interaction (HCI) at PARC and elsewhere has been the development of technologies that *augment human intelligence*. This mission originates with Douglas Engelbart, who inspired researchers like Alan Kay at PARC in the development of the personal computer. The aim of augmented human cognition has remained a core value in the development of, for example, information visualizations, information foraging theory, personalized search, and information scent tools and technologies.

Over the last few years, we have realized that many of the information environments are gradually turning people into social foragers and sharers. People spend much time in communities, and they are using these communities to share information with others, to communicate, to commiserate, and to establish bonds. This is the "Social Web". While not all is new, this style of enhanced collaboration is having an impact on people's online lives.

Augmented Social Cognition research area at PARC has emerged from this background of activities aimed at understanding and developing technologies that enhance the intelligence of users, individually and in social collectives, through socially mediated information production and use. In part this is a natural evolution from our work around improving information seeking and sense making on the Web. In part this is also a natural expansion in our scientific efforts to understand and enhance the intelligence of the individual users coupled to information systems.

Research in Augmented Social Cognition is aimed at enhancing the ability of a group of people to remember, think, and reason; to augment their speed and capacity to acquire, produce, communicate, and use knowledge; and to advance collective and individual intelligence in socially mediated information environments.

In this paper, we describe the emergence of this research endeavor, and summarize some results from the research. For example, we have found that (1) analyses of conflicts and coordination in Wikipedia have shown us the scientific need to understand social sensemaking environments; and (2) information theoretic analyses of social tagging behavior in delicious shows the need to understand human vocabulary systems. We also examine a prototype system in which we explore (3) how decreasing interaction costs might change participation in social tagging systems.

2. WHAT IS AUGMENTED SOCIAL COGNITION?

A natural extension of augmenting human intelligence in the Social Web and Web2.0 world is the development of technologies that augment social intelligence. In this spirit, the meaning of "Augmented Social Cognition" builds on Engelbart's vision, and can be explained by deconstruction of the term:

- ***Cognition*** means the ability to remember, think, and reason; the faculty of knowing; to have functions associated with intelligent action such as perceiving, remembering, planning, deliberating, and learning (acquiring knowledge and experience).
- ***Social Cognition***[1] is the ability of a group of people to remember, think, and reason; the construction of knowledge structures by a group of people; socially mediated acquisition and use of knowledge.
- ***Augmented Social Cognition*** means the enhancement via technical systems of the ability of a group of people to remember, think and reason, acquire and use knowledge.

Technology Trends

Our interest in this area obviously also arouse due to the emergence of Web2.0 and Social Web applications. Web2.0 is a broad term used to mean a new wave of new technologies that is hitting the Web in full-force. What is different about this new Web 2.0 environment is that people are sharing information today in a fundamentally different way from how they are used to. One example is Wikipedia, which is a fascinating collaborative editing environment for creating an encyclopedia. Another example is the various social tagging systems, such as the photo-sharing site flickr.com and URL-sharing site delicious.com.

This wave of new technologies is generated by a combination of new developments, including:

1. *Software as a Service or Web as Platform*. Web technologies have advanced to the point that the Web itself (and other connected networks) has become a computing platform for the delivery of novel features, tools, applications, and services. The computing platform involves a heterogeneous mix of technologies including REST, XML Web Services, RSS/Atom, and AJAX. The web platform provides the plumbing and necessary parts to support rich user interaction, mashups or remixing of Web Services, and the formation of social groups and interactions. *Mashups*: One consequence of the Web as platform is that it fosters innovative combinations of services, such as the connection of search engines or RSS feeds to Google Maps (a web service) to deliver results on geographical data to the end-user.
2. *Rich interaction*. New Web user interfaces no longer rely on the old paradigm of submitting results to the server and waiting for a new page to load. Instead, in its place, we have rich interactive applications that use asynchronous communication to servers to deliver fully interactive user experiences. With higher bandwidth, not only is there more "rich media" (e.g., video), but a richer variety of user-friendly ways to interact with content.
3. *Harnessing network effects of knowledge production and use*. Perhaps the most significant and exciting consequence of the evolution in technology is the emergence of novel architectures of participation that draw users to contribute value, and that gain value as more users cooperate. Novel systems support the creation and aggregation of knowledge through cooperative peer production (e.g., Wikis, blogs, social bookmarking), and others that augment intelligence through cooperative reasoning and judgment (e.g., prediction markets; voting).

Research trends

Researchers are also similarly seeing a surge of new research on Web2.0 technologies distributed in a wide variety of disciplines and associated conferences.

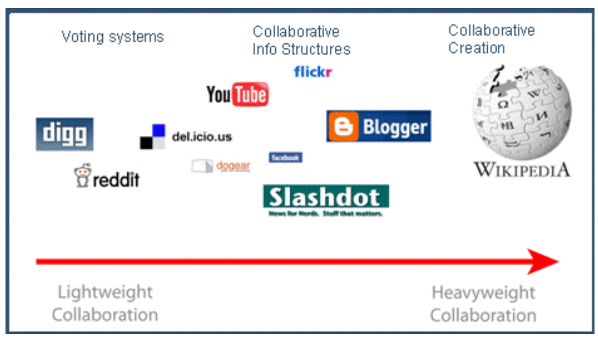


Figure 1: research spectrum in Augmented Social Cognition.

- At the light-end of collaboration spectrum, we have researchers trying to understand the micro-economics of voting systems, of individual and social information foraging behaviors, processes that govern information cascade, and wisdom-of-the-crowd effects. Economists are trying to understand peer production systems, new business models, and consumption and production markets based on intrinsic motivations.
- At the middle of the collaboration spectrum, researchers are building algorithms that mine new socially constructed knowledge structures and social networks. Here physicists and social scientists are using network theories and algorithms to model, mine, and understand these processes. Algorithms for identifying expertise and information brokers are being devised and tested by information scientists.
- At the heavy-end of the collaboration spectrum, the understanding of coordination and conflict costs are especially important for collaborative creation systems such as Wikis. Researchers had studied characteristics that enable groups of people to solve problems together

[1] "Social cognition" has been used for years in psychology to designate the cognitive mechanisms people employ in social interactions. (See for example, Z. Kunda, *Social Cognition: Making Sense of People*, MIT Press, 1999.) Our definition here is intended to include this previous definition, as cognition around social interactions is often a component of the social construction of knowledge structures.

or collaborate on scientific endeavors. Discoveries such as the identification of invisible colleges have shown that implicit coordination can be studied and characterized.

Also, modelers and scientists are trying to understand how to bring down the cost of social interactions, and understand the cost/reward structure for individuals. They are also building characterization models of what, how, and why people are behaving the way they do. Field studies, log file and content analysis, as well as cognitive task analysis are possible studies to conduct in this space.

3. APPLYING SCIENTIFIC RESEARCH METHODS TO AUGMENTING SOCIAL COGNITION

One way to do scientific research on the Social Web is to engage with real users in 'Living Laboratories', in which researchers either adopt or create real useful systems that are used in real settings that are ecologically valid.

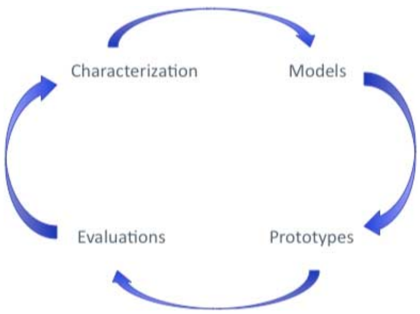


Figure 2: A way to think about the role of Living Laboratory prototypes in scientific research.

This enables a tight loop between characterization of behavior, models of the users and system, prototype, and experimentation. For prototyping, the new Social Web platform is enabling researchers to build systems with amazing speed, enabling the whole loop to be completed within much shorter amounts of time than the past. Ways of looking at real data and analytical/experimental methods are inseparable from the kinds of science and models that can be build in a field.

Here we will look at example research results from each one of these stages of research from Augmented Social Cognition.

4. CHARACTERIZATIONS AND MODELING OF SOCIAL SYSTEM BEHAVIOR

The first step in any new research endeavor is to take out a big piece of paper, and with the aid of large data analytics capabilities provided by databases and MapReduce systems such as Hadoop, to plot and understand the various characteristics of the data.

Here we illustrate our efforts in these areas by looking at the entire revision history of large collaborative systems such as Wikipedia (currently over 8 terabytes of revision history data), and delicious (currently over 500 million bookmarks). Though the data here reported were a bit older, but the both data set was still substantial. For example, the Wikipedia data analyzed in our 2007 paper [Kittur07] contained 50 million+ revisions and around 1 terabytes of data analyzed using Hadoop (hadoop.apache.org).

4.1 Wikipedia Behavior Characterizations: Modeling Wikipedia Growth and Conflicts

As an example of building models and understanding how Web2.0 systems operate, we have been engaged in understanding how conflicts and coordination works in Wikipedia [Kittur07]. Wikipedia, a wiki-based encyclopedia, has become one of the most successful experiments in collaborative knowledge building on the Internet. As Wikipedia continues to grow, the potential for conflict and the need for coordination increase as well. Researchers have seen similar costs in other computer mediated communication (CMC) systems such as MOOs and MUDs [Curtis92, Dibbell93]. Even though researchers have documented the growth of Wikipedia [Voss05], the impact of coordination costs has largely been ignored. Conflict in online communities is a complex phenomenon. Though often viewed in a negative context, it can also lead to positive benefits such as resolving disagreements, establishing consensus, clarifying issues, and strengthening common values [Franco95].

4.1.1 Global Coordination

Here we try to understand the conflict and coordination costs through the concept of indirect work. Viewed from the goal of trying to create high quality content for a collaborative encyclopedia, we define "indirect work" or "conflict and coordination costs" as *excess work in the system that does not directly lead to new article content.* This allows us to develop quantitative measures of coordination costs, and also has broader implications for systems in which maintenance and consolidation occur.

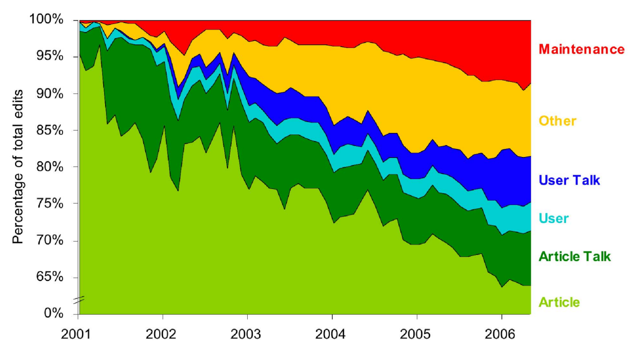


Figure 3. Changing percentage of edits over time showing that decreasing direct work (article) and increasing indirect work (article talk, user, user talk, other, and maintenance).

Overall, user, user talk, procedure, and other non-article pages have become a larger percentage of the total edits made in the system. These trends are summarized in Figure 3, which clearly shows the decreasing percentage of edits going to direct work (article edits) and the increasing percentage of edits going to indirect work across different page types.

4.1.2 Article Conflicts

We wanted to better understand and characterize article-level conflicts. Our goal was to develop an automated way to identify what properties make an article high in conflict using machine learning techniques and simple, efficiently computable metrics. We used the Support Vector Machine (SVM) algorithm to learn what page features predict article conflict scores.

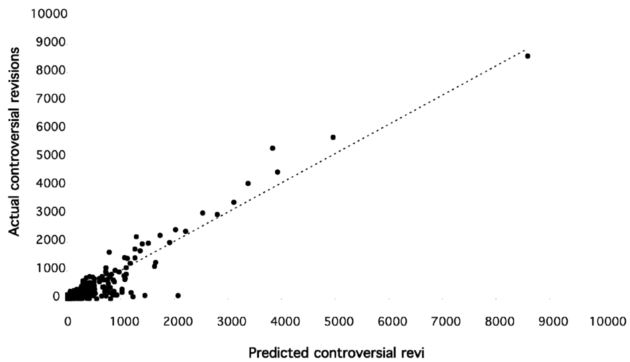


Figure 4. Model performance on articles tagged as controversial. R^2 = 0.897.

The machine learner provides insight to this in the weights it assigns to various page metrics. These weights are determined by the utility of a metric in predicting CRC scores, and are shown in order of importance in

Table 1.

Table 1. Highly weighted metrics, rank ordered. Up arrows indicate positive correlation with conflict; down arrows indicate negative correlation with conflict

↑	1. Revisions (article talk)
↑	2. Minor edits (article talk)
↓	3. Unique editors (article talk)
↑	4. Revisions (article)
↓	5. Unique editors (article)
↑	6. Anonymous edits (article talk)
↓	7. Anonymous edits (article)

By far the most important metric to the model was the number of revisions made to an article talk page (#1 above). This is not unexpected, as article talk pages are intended as places to discuss and resolve conflicts and coordinate changes. Some of the metrics are more surprising; for example, one might expect that the more points of view are involved, the more likely conflicts will arise. However, the number of unique editors involved in an article *negatively* correlates with conflict (#5 above), suggesting that having more points of view can defuse conflict.

Another interesting finding is that while anonymous edits to the article talk page correlate with increased conflict (#6), they correlate with reduced conflict when made to the main article page (#7). This suggests that anonymous editors may be valuable contributors to Wikipedia on the article page where they are adding or refining article content. However, anonymity on the article talk page, where heated discussions often occur, seems to fan the flames. This suggests that anonymity may be a two-edged sword, useful in lowering participation costs for content but less so in conflict resolution situations.

4.1.3 User Conflicts

The characterization of conflicts between users is crucial to understanding the motivation of users and the sources of conflicts. The goals are to 1) identify users involved in conflicts; 2) characterize ongoing conflicts; and 3) develop a tool that can help in understanding the conflicts.

We built a tool called Revert Graph to visualize user conflict on a particular article. Revert Graph retrieves all users who have participated in reverts and visualizes a graph based on revert relationships between the users (Figure 5 and Figure 6).

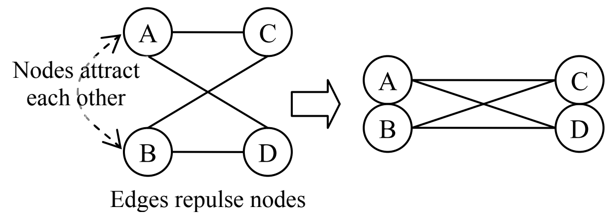


Figure 5. Force directed layout structure employed in Revert Graph. Users (represented as nodes) attract each other unless they have a revert relationship. A revert is represented as an edge. When there are reverts between users, they push against each other. Left figure: Nodes are evenly distributed as an initial layout. Right figure: When forces are deployed, nodes are rearranged in two user groups.

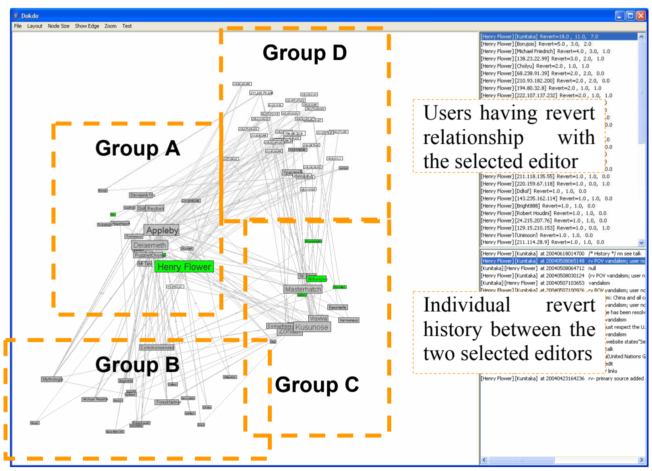


Figure 6. Revert Graph for the Wikipedia page on *Dokdo*. Revert Graph uses force directed layout to simulate revert relationship between users. The tool also allows users to drill down into revert relationships, which enables them to investigate the nature of the conflicts.

We can identify user clusters based on the assumption that a group of users have closer views on a topic the more they revert users in another user group.

The Wikipedia page on Dokdo (Figure 6) is one example where we were able to find interesting user clusters. Dokdo is a disputed islet in the Sea of Japan (East Sea) currently controlled by South Korea, but also claimed by Japan as Takeshima. Figure 6 shows user groups discovered on the Dokdo article. We manually labeled each user based on his/her position on the issue. The majority of users in Group A supports the Korean claims while users in Group C show the opposite pattern. Unlike Group A and C, users in Group D and B showed mixed opinion on the issue.

4.2 Delicious Behavior Characterizations: Modeling Social Tagging Vocabulary using Information Theory

Given the rise in popularity of social tagging systems, it seems only natural to ask how efficient is the organically evolved tagging vocabulary in describing any underlying document objects? Does this distributed process really provide a way to circumnavigate the traditional categorization problem with ontologies? Shirky argues that since tagging systems does not use a controlled vocabulary, it can easily respond to changes in the consensus of how things should be classified [Shirky05].

Furnas mentioned that a potential cognitive process for explaining how social tagging works might arise out of an analysis of the "vocabulary problem" [Furnas06]. Specifically, Furnas mentioned that the process for generating a tag for an item that might be needed later appears to be the same process that is used to generate search keywords to retrieve a particular item in a search and retrieval engine.

Furnas' comment pointed to the usefulness of social tagging systems as a communication device that can bridge the gap between document collections and users' mental maps of those collections. Social navigation as enabled by social tagging systems can be studied by how well the tags form a vocabulary to describe the contents being tagged.

We analyzed a social tagging site, namely delicious.com, with information theory in order to evaluate the efficiency of this social tagging site for encoding navigation paths to information sources [Chi07].

We show that entropy analysis from information theory provides a natural way to understand the descriptive encoding power of tags, which appears to be weaning. We found that users appear to have responded by increasing the number of tags they use to describe each item. This metric should be helpful in future analysis of social tagging systems.

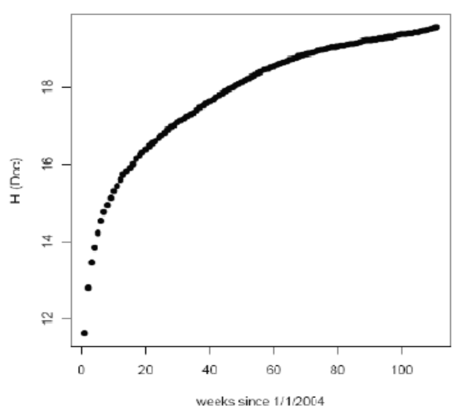


Figure 7. Entropy of documents H(D) is increasing over time.

As shown in Figure 7, one can see that the entropy of the document set, *H(D)*, continued to increase. We know that the number of documents in the system is increasing, contributing to this increase in entropy. This means that, over time, users continue to introduce a wide variety of new documents into the system and that the diversity of documents is increasing over time.

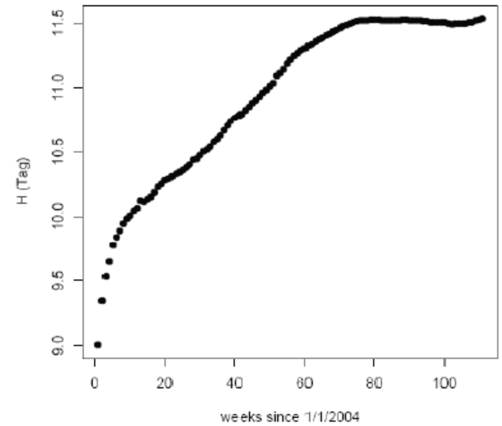


Figure 8. Entropy of tags *H(T)* is increasing at first, then started to plateau around Week 75 (mid-2005).

Figure 8 shows a marked increase in the entropy of the tag distribution *H(T)* up until week 75 (mid-2005) at which point the entropy measure hits a plateau. At the same time, the total number of tags is increasing, even during the plateau section. Since the total number of tags kept increasing, tag entropy can only stay constant in the plateau by having the tag probability distribution become less uniform. What this suggests is that eventually the tagging vocabulary saturated, and coming up with new keywords became difficult. That is to say, a user is more likely to add a tag that is already popular than to add a tag that is relatively obscure.

More importantly, the entropy of documents conditional on tags, *H(D|T)*, is increasing rapidly (see Figure 9). What this means is that, even after knowing completely the value of tags, the entropy of the document is still increasing. Conditional Entropy gives us a method for analyzing how useful a set of tags is at describing a document set. The fact that this curve is strictly increasing suggests that the specificity of any given tag is decreasing. That is to say, as a navigation aid, tags are becoming harder and harder to use. We are moving closer and closer to the proverbial "needle in a haystack" where any single tag references too many documents to be considered useful.

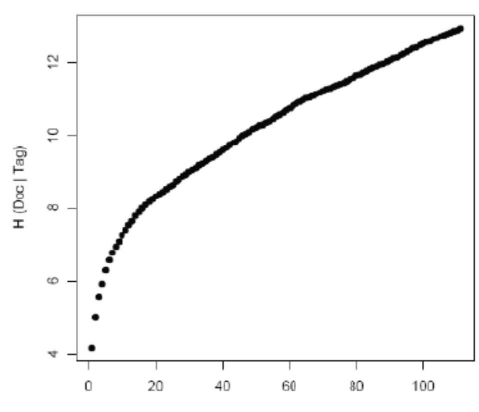


Figure 9. Entropy of Documents conditional on Tags *H(D|T)* increases over time.

4.3 Social Search Behavior Modeling: Understanding Social Search Using Mechanical Turk Surveys

Information retrieval researchers typically depict information seeking as solitary activities of a single person in front of a web browser. This view is slowly changing.

Researchers and practitioners now use the term "social search" to describe search systems in which social interactions or information from social sources is engaged in some way [Evans and Chi 2008]. These recent trends point to the social nature of information seeking. Indeed, we recently conducted research with 150 participants using Mechanical Turk surveys [Kittur08], which suggested that many information-seeking activities are interwoven in-between social interactions [Evans and Chi, 2008]. Our research suggests analyzing the search process by looking at three stages of before, during, and after the search:

Before: We saw users engaged in social interactions 43% of the time before exploring on the web. These social interactions supported information gathering by providing opinions and advice such as websites or keywords to try. For example, a programmer might have engaged in a series of emails with coworkers asking about the existence of various monitoring software packages for a web server, and the merits of each package. An analysis of only search engine logs might have simply recorded several refinements of queries in a single 30-minute session rather than detecting the considerable amount of social preparation done before searches.

During: Social interactions are also common during the search act itself. For example, people sometimes searched with others who are co-located, in which they might take turns suggesting and trying out search keywords. In these cases, users are likely to interact with others during informational exploratory searches. Around 40% of the users engaged with others both before and during the information search.

After: Users often either organize the search results or distribute them to others in their social network. For example, after a barista found a particular recipe, she printed it out and shared it with all of her coworkers. In fact, we observed users distribute search results to others quite frequently at around 60%.

We have integrated our findings with models from previous work on sensemaking and information-seeking behaviors [Evans and Card, 2008] to present a canonical model of social search. Figure 1 below depicts this descriptive model. We see that, when viewed holistically, information seeking is more than just a database query. Instead, information seeking is often embedded within social relationships. The social networks are both sources of requests as well as suggestions. They are also sinks in which refined results are distributed.

Our results and analysis demonstrated that users have a strong social inclination throughout the search process, interacting with others for reasons ranging from obligation to curiosity. Self-motivated searchers and users conducting informational searches provided the most compelling cases for social support during search.

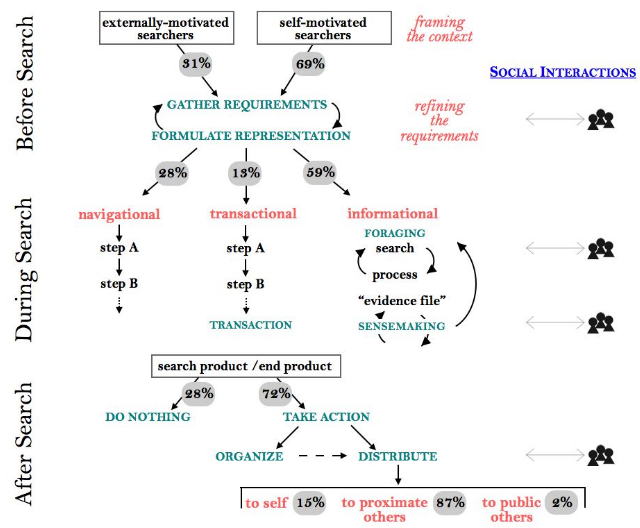


Figure 10. Combining with previous models of information seeking behavior, a canonical model of social search shows three stages in a search act weaved in between social interactions.

Current social search systems can be categorized into two general classes:

Social answering systems utilize people with expertise or opinions to answer particular questions in a domain. Answerers could come from various levels of social proximity, including close friends and coworkers as well as the greater public. Yahoo! Answers (answers.yahoo.com) is one example of such systems. Early academic research includes Ackerman's Answer Garden [Ackerman, 1996], and recent startups include Mechanical Zoo's Aardvark (vark.com) and ChaCha's mobile search (chacha.com).

Some systems utilize social networks to find friends or friends of friends to provide answers. Web users also use discussion forums, IM chat systems, or their favorite social networking systems like Facebook and Friendfeed to ask their social network for answers that are hard to find using traditional keyword-based systems. These systems differ in terms of their immediacy, size of the network, as well as support for expert finding.

Importantly, the effectiveness of these systems depends on the efficiency in which they utilize search and recommendation algorithms to return the most relevant past answers, allowing for better constructions of the knowledge base.

Social feedback systems utilize social attention data to rank search results or information items. Feedback from users could be obtained either implicitly or explicitly. For example, social attention data could come from usage logs implicitly, or systems could explicitly ask users for votes, tags, and bookmarks. Direct Hit[2] was one early example from early 2001 that used click data on search results to inform search ranking. The click data was gathered implicitly through the usage log. Others like Wikia Search (search.wikia.com), and most recently Google, are allowing users to explicitly vote for search results to directly influence the search rankings.

[2] http://www.searchengineshowdown.com/features/directhit/review.html

Vote-based systems are becoming more and more popular recently. Google's original ranking algorithm PageRank could also be classified as an implicit voting system by essentially treating a hyperlink as a vote for the linked content. Social bookmarking systems such as del.icio.us allow users to search their entire database for websites that match particular popular tags.

One problem with social cues is that the feedback given by people is inherently noisy. Finding patterns within such data becomes more and more difficult as the data size grows [Chi and Mytkowicz, 2008]

In both classes, there remains opportunity to apply more sophisticated statistical and structure-based analytics to improve search experience for social searchers. For example, expertise-finding algorithms could be applied to help find answerers who can provide higher-quality answers in social answering systems. Common patterns between question-and-answer pairs could be exploited to construct semantic relationships that could be used to construct inferences in question answering systems. Data mining algorithms could construct ontologies that are useful for browsing through the tags and bookmarked documents.

5. PROTOTYPING REAL SYSTEMS

We have been building real Social Web systems and releasing them into the real world in 'Living Laboratory' fashion. Systems described here can all be found on the Web running on real world data sets.

5.1 WikiDashboard

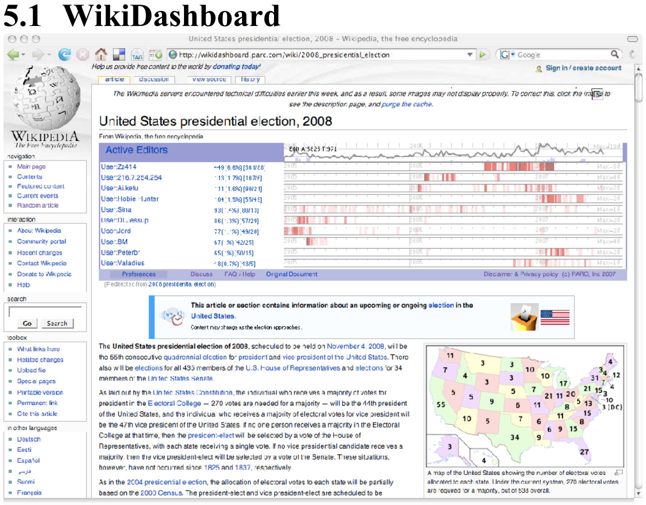


Figure 11. WikiDashboard is a visualization overlay for live Wikipedia pages. The dashboard provides a useful visual digest about who edits how many revisions on each Wikipedia page. It allows users to easily evaluate social activities and patterns around the page, which may be hard to detect otherwise. This figure shows an example of the tool applied to the Wikipedia article "United States presidential election, 2008"

Accountability has been recognized as an important factor influencing trust in many online interactions and it plays an increasingly important role in collaborative knowledge systems such as wikis [Denning05]. Although users can access past revisions of every page, it is difficult and time-consuming even for dedicated users to make sense of the history of a page, because many page histories run into the thousands of edits. We are investigating how providing access to this type of accountability information, i.e. who edits how many revisions for an article, in a digestible form could affect users' trust and interpretation of an article. If so, the approach can result in reducing the risks many perceive as inherent to a system [Denning05] in which anyone can contribute or change anything.

To address this challenge, we designed WikiDashboard (http://wikidashboard.parc.com), a tool that helps users to identify interesting edit patterns in Wikipedia pages, patterns that may be very hard to detect otherwise [Suh07]. As shown in Figure 11, the site provides a dashboard for each page in Wikipedia, while proxying the rest of the content from Wikipedia. The dashboard provides a visualization overlay onto every live Wikipedia page, enabling users to be aware of social dynamics and context around the page they are about to read. The prototype can be used just as if users are on the Wikipedia site itself.

Each article has an associated article dashboard that displays an aggregate edit activity graph representing the weekly edit trend of the article, followed by a list of the active users who made edits on the page.

A user page is like a home page to display information relating to a user. In our system, each user page has a User Dashboard embedded, displaying the article contribution and editing patterns of that user (Figure 12).

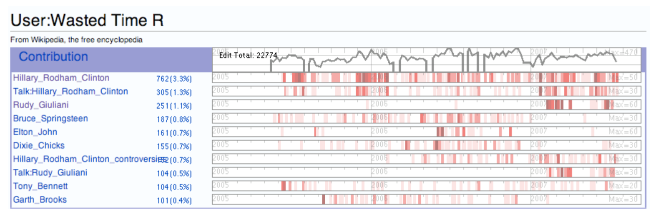


Figure 12. User Dashboard is embedded in each user page of Wikipedia. The dashboard displays weekly edit trend of an editor as well as the list of articles that the editor made revisions on. This example shows a user, "Wasted Time R" made significant edits on articles related to New York politicians and pop singers.

Theories of social translucence [Erickson02] state that three building blocks are necessary for effective communication and collaboration: making socially significant information visible and salient; supporting awareness of the rules and constraints governing the system; and supporting accountability for actions. The idea of social translucence suggests that WikiDashboard could benefit not only readers but also improve the effectiveness of active writers.

5.2 MrTaggy: a social search browser based on social tagging data

At PARC, we have been constructing a social search system based on statistical machine learning. Our system, called MrTaggy (mrtaggy.com), relies on 150 million bookmarks crawled from the web to construct a similarity graph between tag keywords. MrTaggy's tag-search browser uses this similarity graph to recommend and search through other tags and documents.

Figure 13. MrTaggy's user interface with related tags list on the left and search results lists presented on the right.

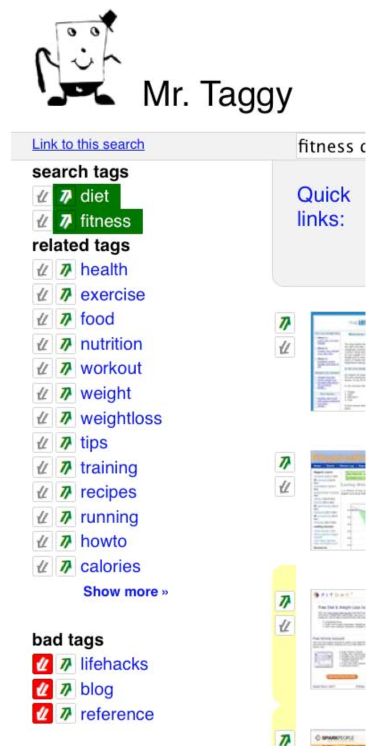


Figure 14. MrTaggy user interface for adding relevant and irrelevant tags.

The Figure above shows a typical view of the tag search browser. MrTaggy provides typical search capabilities (query input textbox and search results list) combined with explicit relevance feedback for query refinements. Users have the opportunity to give relevance feedback to the system in two different ways, at the fine-grained item level and at a coarse descriptor (tag) level:

Related Page Feedback: Clicking on the upward or downward arrow on a search result includes or excludes it from the result list. This feedback also results in emphasis of other similar or de-emphasis of other dissimilar web pages.

Related Tag Feedback: On the left a *related tags list* is presented, which is an overview of other tags related to the current set of tag keywords. For each related tag, up and down arrows are displayed to enable the user to give relevance feedback by specifying relevant or irrelevant tags.

For a search result, MrTaggy displays the most commonly used tags describes the content of the web page, in addition to the title and the URL of the corresponding web page. Other people applied these tags to label the corresponding Web page. When hovering over tags presented in the snippet, up and down arrows are displayed to enable relevance feedback on these tags as well.

Having just described the interaction of the relevance feedback part of the system, we now describe how it operates in concert with the backend. Figure 15 below shows an architecture diagram of the overall system.

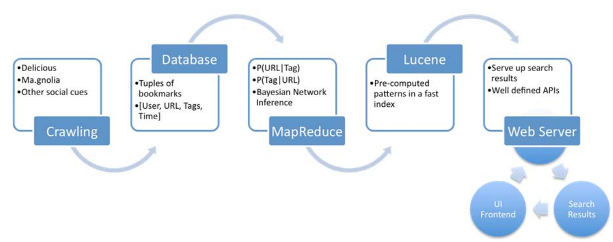


Figure 15. Overall architectural diagram of the MrTaggy tag-based search browser.

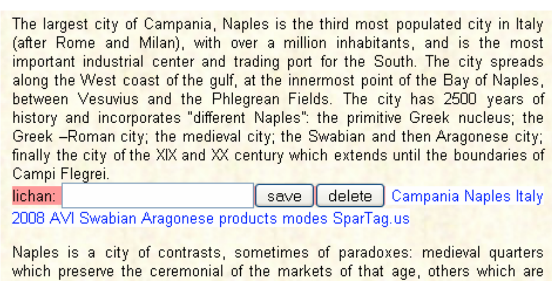
First, a crawling module goes out to the web and crawls social tagging sites, looking for tuples of the form <*User, URL, Tag, Time*>. Tuples are stored in a MySQL database. In our current system, we have roughly 150 million tuples. A MapReduce system based on Bayesian inference and spreading activation then computes the probability of each URL or tag being relevant given a particular combination of other tags and URLs. Here we first construct a bigraph between URLs and tags based on the tuples and then precompute spreading activation patterns across the graph. To do this backend computation in massively parallel way, we used the MapReduce framework provided by Hadoop (hadoop.apache org). The results are stored in a Lucene index (lucene.apache.org) so that we can make the retrieval of spreading activation patterns as fast as possible.

Finally, a web server serves up the search results through an interactive frontend. The frontend responds to user interaction with relevance feedback arrows by communicating with the web server using AJAX techniques and animating the interface to an updated state.

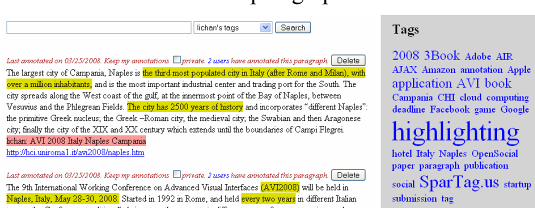
5.3 SparTag.us: A Low Cost Paragraph-based Tagging System for Foraging of Web Content

Tagging systems such as del.icio.us and Diigo have become important ways for users to organize information gathered from the Web. However, despite their popularity among early adopters, tagging still incurs a relatively high interaction cost for the general users.

Information gathering and sharing are essential steps towards the goal of social sensemaking. In the past few years, a variety of Web 2.0 tools have been introduced to support social information foraging and sensemaking. In order to understand how social cognition can be augmented, we must understand the individuals' incentive to contribute content to the larger social group. SparTag.us dramatically lowers the cost of interaction to try to understand whether lowering the cost of participation increases participation productively.



(a) Clicking on the paragraph inserts a tagging widget to the end of the paragraph.



(b) The top portion of the reading notebook that SparTag.us created for user *lichan***.**

Figure 16: SparTag.us system features include click2tag and paragraph fingerprinting.

We introduced a new tagging system called SparTag.us [Hong08], which uses an intuitive Click2Tag technique to provide in situ, low cost tagging of web content. SparTag.us also lets users highlight text snippets and automatically collects tagged or highlighted paragraphs into a system-created notebook, which can be later browsed and searched. Motivated by the prominence of redundant contents on the Web with different URLs and shared documents that are read and re-read within enterprises, we explore the idea of paragraph fingerprinting to achieve the goal of "annotate once, appear anywhere" [Hong09].

6. EVALUTION

Having released these systems, we want to find out how well they would really work with real users. Ideally, evaluations of these systems would occur after there are substantial adoptions and usage of the systems in the real world. In particular, WikiDashboard has been available for around for around 1.5 years with over 50,200 visits and 168,300 page views. Thus, we have already been able to capture a number of insightful feedbacks from various users:

"*WikiDashboard appears to be a valuable tool that can provide some good insights into individual edit patterns and edit conflicts on specific articles. As a means of learning about the tool I have found it useful to use it on articles that I have an intimate understanding of development in order to get a feel of how it can be used and interpreted.*" [3]

"*This is very useful for getting a quick glance of the user's editing interests over time. ... I actually think a tool like WikiDashboard presents significantly more utility, and is the beginning of an interesting trend of repurposing metadata to create a trust heuristic*." [4]

However, for many systems, we perform a laboratory study before we have captured enough real-world usage. Here we report on some examples of these types of evaluation.

6.1 WikiDashboard Study

We recently reported a user study conducted using Amazon's Mechanical Turk showing how dashboards affects user's perception of trustworthiness in Wikipedia articles [Kittur08].

In that experiment, we designed nearly identical dashboards in which only a few elements are changed. We designed a visualization of the history information of Wikipedia articles that aggregates a number of trust-relevant metrics.

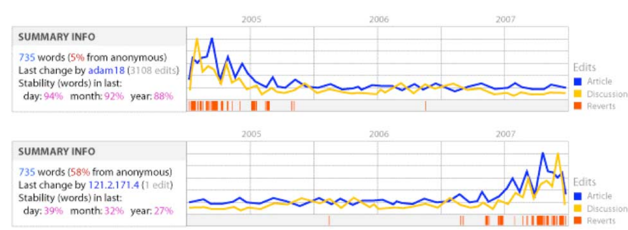


Figure 17: high-trust and low-trust versions developed for WikiDashboard experiment.

We developed high-trust and low-trust versions of the visualization by manipulating the following metrics:

- Percentage of words contributed by anonymous users. Anonymous users with low edit-counts often spam and commit vandalism.
- Whether the last edit was made by an anonymous user or by an established user with a large number of prior edits.

---

[3] The Wikipedia Review, http://www.wikipediareview.com

[4] Unit Structures, http://chimprawk.blogspot.com/

- Stability of the content (measured by changed words) in the last day, month, and year.
- Past editing activity. Displayed in graphical form were the number of article edits (blue), number of edits made to the discussion page of the article (yellow), and the number of reverts made to either page (red). Each graph was a mirror image of the other, and showed either early high stability with more recent low stability, or vice versa.

We also included a baseline condition, in which no visualization is used at all.

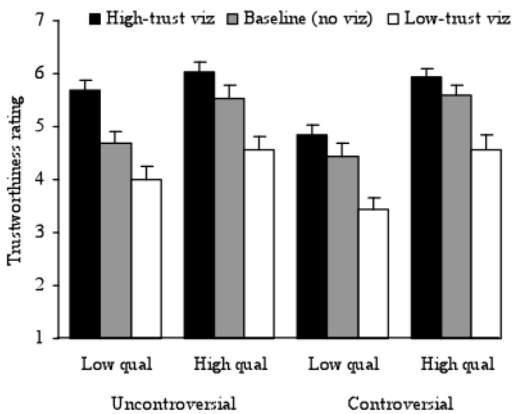


Figure 18: Results from the WikiDashboard evaluation.

The results with Mechanical Turk users show that surfacing trust-relevant information had a dramatic impact on users' perceived trustworthiness, holding constant the content itself. The effect was robust and unaffected by the quality and degree of controversy of the page. Trust could be impacted both positively and negatively. High-trust condition increased trustworthiness above baseline and low-trust condition decreased it below baseline. This result is obviously very encouraging for folks who are keeping score on the effects of *transparency* on trust.

These results suggest that the widespread distrust of wikis and other mutable social collaborative systems may be reduced by providing users with transparency into the stability of content and the history of contributors.

6.2 MrTaggy Searching and Browsing Study

We recently completed a 30-subject study of MrTaggy [Kammerer et al. 2008]. In this study, we analyzed the interaction and UI design. The main aim was to understand whether and how MrTaggy is beneficial for domain learning.

We compared the full exploratory MrTaggy interface to a baseline version of MrTaggy that only supported traditional query-based search. We tested participants' performance in three different topic domains and three different task types. The results show:

(1) Subjects using the MrTaggy full exploratory interface took advantage of the additional features provided by relevance feedback, without giving up their usual manual query typing behavior. They also spent more time on task and appear to be more engaged in exploration than the participants using the baseline system.

(2) For learning outcomes, subjects using the full exploratory system generally wrote summaries of higher quality compared to baseline system users.

(3) To also gauge learning outcomes, we asked subjects to generate keywords and input as many keywords as possible that were relevant to the topic domain in a certain time limit. Subjects using the exploratory system were able to generate more reasonable keywords than the baseline system users for topic domains of medium and high ambiguity, but not for the low-ambiguity domain.

Our findings regarding the use of our exploratory tag search system are promising. The empirical results show that subjects can effectively use data generated by social tagging as "navigational advice". The tag-search browser has been shown to support users in their exploratory search process. Users' learning and investigation activities are fostered by both relevance feedback mechanisms as well as related tag ontologies that give scaffolding support to domain understanding. The experimental results suggest that users' explorations in unfamiliar topic areas are supported by the domain keyword recommendations presented in the related tags list and the opportunity for relevance feedback.

6.3 SparTag.us Social Reading Study

We conducted a 'Social Reading Experiment' where participants needed to use Web resources to learn about a topic area: "Enterprise 2.0 Mashups", which is a combination of the technology areas of "Enterprise 2.0"[5] and "Web 2.0 Mashups". Study participants would need to find and understand many web pages because at the time of the study there was no single source of information on the topic area. Our experiment compared three groups of participants who worked:

1) Without SparTag.us (WS), but with traditional note-taking tools.
2) With SparTag.us only, used individually (SO).
3) With SparTag.us with the annotations of a 'Friend' (SF).

The conditions WS and SO were control conditions in which individuals read web content without access to others' annotations. To provide for an ecologically valid comparison, WS participants could take notes in MS Word or with pen and paper. In the SF condition, people independently read web content but also had access to social annotations created by an experimenter-simulated subject-matter expert.

Tools like SparTag.us and del.icio.us are tools used at an Internet scale and scope. In our experimental setup we look at the performance of individual users. However, we extended the scope of inquiry beyond the individual by simulating a social reading condition. That is, in one of the conditions each user was exposed to the SparTag.us Friend, which is an organized collection of annotations comprising a tag cloud, a list of URLs, and a set of paragraphs.

The hypothesis is that participants that were exposed to tags, URLs, and highlights from a knowledgeable other would perform better than the participants without this exposure. We thus evaluate performance measures between subjects in the

[5] http://en.wikipedia.org/wiki/Enterprise_2.0

experimental condition, SF, with those in control conditions, SO and WS. Eighteen participants completed two experimental sessions. The first day was a four hour series of demographic survey, true-false question answering, learning in the domain area lasting two hours, one writing essay, and a debrief. Day 2 lasted one hour and involved one true-false question set and a second writing task. More details on the procedure can be found in [6].

We used a combination of performance and process measures to understand the impact of the annotation support used, but also give indications of how people are employing the technology in the context of their reading and annotation practices. The performance was measured using a questionnaire (created for this study). The questionnaire included a set of true-false questions, which were generated from an expert elicitation process and were used to assess objective learning gains in the subject matter domain before and after the users foraged the information in each of the three conditions.

The process measures pertained to the reading and writing behaviors of each participant: the number and sequence of Web resources visited (logged by Universal Resource Locator or URL), loaded and scrolled; the annotations made (tags and keywords used), and the personal notes taken during the task.

The main measure of learning (equation (1)) was obtained through a metric of learning effect developed as part of the experimental method. The Gain metric is a composite indicator that was computed on the basis of several scores derived from the questionnaire: Pretest to Posttest questionnaire scores for each participant, and maximum score. Specifically, gain scores were calculated as:

$$Gain = \frac{(PostTest\_Score - PreTest\_Score)}{(Max\_Score - PreTest\_Score)} \qquad \textbf{(1)}$$

Using the Gain metric as the measure of learning performance, we report in [Nelson09] a learning effect, with the SF group showing significantly greater gains than the SO group and the WS group. The WS and SO groups were not significantly different.

The mean gain scores were: SF group, M=0.46, SD=0.22; SO group, M=0.13, SD=0.32; WS group, M=0.27, SD=0.23. An analysis of covariance showed a significant effect of learning group, $F(2, 16) = 5.91$, $p < .05$, with the SF group showing significantly greater gains than the SO group, $t(16) = 4.66$, $p < .0005$, and the WS group, $t(16) = 3.93$, $p = .001$. The WS and SO groups were not significantly different.

This establishes that participants with access to resources from a knowledgeable other exhibited a greater learning performance.

7. CONCLUSION

Augmented Social Cognition is a new area to understand and develop engineering models for systems that enhance a group's ability to remember, think, and reason. While more enterprises contemplate the benefits of Web 2.0 social software (enhanced collaboration, innovation, knowledge sharing), the coordination and interaction costs that occur in social systems are often overlooked. In this article, we outlined our recent research:

First, we are characterizing and modeling the various social web spaces in order to understand its collaboration and coordination models. Based on extensive studies of social systems such as delicious and Wikipedia, we have started to identify multiple factors that must be managed to realize the full benefits of these systems.

Second, we are building new social web applications based on the concepts of social transparency and balancing interaction costs and participation levels. We are then evaluating these web applications to understand whether they have the capacity to really improve social systems.

8. ACKNOWLEDGMENTS

This technical note contains summarized research results from collaborations with current and past members of the ASC research group. I thank them for great advice and comments.

9. REFERENCES

[1] Ackerman, M. S.;McDonald, D. W. 1996. Answer Garden 2: merging organizational memory with collaborative help. In Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (Boston, Massachusetts, United States, November 16 - 20, 1996). M. S. Ackerman, Ed. CSCW '96. ACM, New York, NY, 97-105. DOI= http://doi.acm.org/10.1145/240080.240203

[2] Chi, E. H.; Mytkowicz, T. Understanding the efficiency of social tagging systems using information theory. Proceedings of the 19th ACM Conference on Hypertext and Hypermedia; 2008 June 19-21; Pittsburgh, PA. NY: ACM; 2008; 81-88.

[3] Curtis, P. Mudding: Social phenomena in text-based virtual realities. Palo Alto Research Center, 1992.

[4] Denning, P., Horning, J., Parnas, D., and Weinstein, L. "Wikipedia risks", *Comm. ACM,* 48(12), ACM Press (2005), 152-152.

[5] Dibbell, Julian. A Rape in Cyberspace. *Village Voice, 21* (1993), 26-32.

[6] Erickson, T., Halverson, C., Kellogg, W.A., Laff, M., and Wolf, T. "Social translucence: designing social infrastructures that make collective activity visible", *Comm. ACM*, 45 (4), ACM Press (2002), 40-44.

[7] Evans, Brynn, Ed H. Chi. Towards a Model of Understanding Social Search. In Proc. of Computer-Supported Cooperative Work (CSCW), pp. 485-494. ACM Press, 2008.

[8] Evans B. M. and S. K. Card. Augmented information assimilation: Social and algorithmic web aids for the information long tail. In Proc. CHI2008, ACM Press, pages 989–998, 2008.

[9] Franco, V., Piirto, R., Hu, H. Y., Lewenstein, B. V., Underwood, R., and Vidal, N. K. Anatomy of a flame: conflict and community building on the Internet. *Tech. and Society Magazine, IEEE, 14* (1995) 12-21.

[10] Furnas, G. W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., Davis, M., Marlow, C., and Naaman, M. 2006. Why do tagging systems work?. In *Proc. CHI '06 Extended Abstracts*. ACM Press, New York, NY, 36-39.

[11] Hong, Lichan, Ed H. Chi, Raluca Budiu, Peter Pirolli, and Les Nelson. SparTag.us: Low Cost Tagging System for Foraging of Web Content. In Proceedings of the Advanced Visual Interface (AVI2008), pp. 65--72. ACM Press, 2008.

[12] Hong, L. and Chi, E. H. 2009. Annotate once, appear anywhere: collective foraging for snippets of interest using paragraph fingerprinting. In Proceedings of the 27th international Conference on Human Factors in Computing Systems (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 1791-1794.

[13] Kammerer, Y., Nairn, R., Pirolli, P., and Chi, E. H. 2009. 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 (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 625-634.

[14] Kittur, Aniket, Bongwon Suh, Ed H. Chi. Can You Ever Trust a Wiki? Impacting Perceived Trustworthiness in Wikipedia. In Proc. of Computer-Supported Cooperative Work (CSCW), pp. 477-480. ACM Press, 2008. San Diego, CA. Best Note Award.

[15] Kittur, Aniket, Ed H. Chi, Bongwon Suh. Crowdsourcing User Studies With Mechanical Turk. In Proceedings of the ACM Conference on Human-factors in Computing Systems (CHI2008), pp.453-456. ACM Press, 2008. Florence, Italy.

[16] Kittur, A., Bongwon Suh, Bryan Pendleton, Ed H. Chi. He Says, She Says: Conflict and Coordination in Wikipedia. In Proc. of ACM CHI 2007 Conference on Human Factors in Computing Systems, pp. 453--462, April 2007. ACM Press. San Jose, CA.

[17] Marchionini, G. Exploratory search: From finding to understanding. *Communications of the ACM*, *49,* 4 (2006), 41-46.

[18] Nelson, L., Held, C., Pirolli, P., Hong, L., Schiano, D., and Chi, E. H. 2009. With a little help from my friends: examining the impact of social annotations in sensemaking tasks. In Proceedings of the 27th international Conference on Human Factors in Computing Systems (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 1795-1798.

[19] Shirky, Clay. Ontology is Overrated: Categories, Links, and Tags. Blog entry. http://shirky.com/writings/ontology overrated.html (retrieved Sept 21, 2006).

[20] Suh, B., Chi, E, Kittur, A., Pendleton, B. Lifting the Veil: Improving Accountability and Social Transparency in Wikipedia with WikiDashboard. To appear in Proc. CHI2008 Conference on Human Factors in Computing Systems. ACM Press, 2008.

[21] Voss, J. Measuring Wikipedia. In Proc. 10th Intl. Conf. Intl. Soc. for Scientometrics and Informetrics (2005).

[22] White, R.W., Drucker, S.M., Marchionini, M., Hearst, M., schraefel, m.c. Exploratory search and HCI: designing and evaluating interfaces to support exploratory search interaction. *Extended Abstracts CHI 2007*, ACM Press (2007), 2877-2880.

Short Bio:

Ed H. Chi is area manager and senior research scientist at Palo Alto Research Center's Augmented Social Cognition Group. He leads the group in understanding how Web2.0 and Social Computing systems help groups of people to remember, think and reason. Ed completed his three degrees (B.S., M.S., and Ph.D.) in 6.5 years from University of Minnesota, and has been doing research on user interface software systems since 1993. He has been featured and quoted in the press, such as the Economist, Time Magazine, LA Times, and the Associated Press.

With 19 patents and over 50 research articles, his most well-known past project is the study of Information Scent --- understanding how users navigate and understand the Web and information environments. He has also worked on computational molecular biology, ubicomp, and recommendation/search engines. He has won awards for both teaching and research. In his spare time, Ed is an avid Taekwondo martial artist, photographer, and snowboarder.

.