Dissertation Proposal

Contextual Authority Tagging : Expertise Location via Social Labeling

Terrell Russell

University of North Carolina at Chapel Hill
School of Information and Library Science

April 6, 2010 – r715

Table of Contents

List of Figures

List of Tables

Chapter 1

Summary

Today, the Internet has democratized speech at every level. It has made free and open speech more available to everyone but it has not provided us with the requisite filters to disambiguate the signal from all the new noise. For democratic purposes, it is important that everyone have a voice (and an equal vote), but for most other purposes, it is not necessary. For most purposes, it is most helpful to hear the opinions of those who know what they are talking about and who have the most to offer the conversation.

Reliably knowing who the experts are would be the first step of a larger plan to filter the signal from the noise in our Internet-empowered world where everyone can have a bullhorn.

The fundamental issue of expertise location has been faced at a smaller scale within organizations. Knowing what an organization knows about, and who carries that knowledge, is a valuable asset and has been a primary focus of knowledge management for many years. In large part, *knowing who knows what* has come from two places – the individuals who have self-reported their own expertise and from algorithmic derivation from the produced documents and paper trail of doing business.

I think that a valuable third source is being overlooked. I think that people, other than the individual, have interesting insight and knowledge about what the individual knows. I think that their collective human opinion can serve as a reliable indicator of knowledge as well and should be included.

This research will evaluate the ability of a group to know what an individual knows.

Chapter 2

Background

2.1 Overview

Humans can only sense and process so much. Because of this physical limitation, we have sought shortcuts in order to help us sense "more" (Downs, 1957) and to make up for our limited ability to have encyclopedic knowledge of the situations around us (Lupia, 1994). The use of many of these shortcuts is dependent on other people – those around us, those before us, and those far away. Our dependence on others is inefficient in that we do not always know whom to ask or approach for help. Sometimes we waste valuable time and energy looking for the right source of information. We may be able to reduce this waste with some thoughtful sharing and collective reflection. We could benefit greatly by discovering the latent, undocumented knowledge of those around us and bringing it to the surface. We should be able to tap the implicit by making it more explicit (Nonaka, 1991).

This research is an investigation into how a group of people can come to know what it is that its members know. Through simple keyword tagging and cognitive reflection on those tags over time, an individual and a group of his peers may approach a common ground around the topic of his areas of expertise. Better senses of self-awareness, other-awareness, and downstream decision-making may come about because of this information being collected and shared.

This research is primarily focused on tagging data around humans whose granted cognitive authority (Wilson, 1983) to one another changes over time.

2.2 Problem Statement

Knowledge of our surroundings, from an empiricist perspective, comes from our five senses. The things we see and hear, the things we smell and touch and taste, they are all just constrained representations of our environment. We strive to make as much sense of the world as possible, but we are limited by our physical location, our position in time, access to information resources, and by the processing power of our brains (Dervin, 1983). Cognitive load theory (Sweller et al., 1998) tells us that we can only handle so much data coming in at a time.

Because of this constraint, we seek shortcuts, or second-hand information, in order to "see" more, to see beyond what is readily apparent. We seek shortcuts in order to "know" more than what our senses can sense. I think these pieces of second-hand information can be of two distinct types, either basic pieces of simple information, or information that resembles an executive summary. Second-hand information can come from others in the form of basic facts like "it's raining outside" or "it's raining at the beach" – both of which are simple facts but relayed to us by another, rather than collected or sensed on our own. Second-hand information can also come in the form of more summarized or processed information like "our economy is in a recession". This second type of new information could have been determined by one person or synthesized by many, but it also comes from sources outside of ourselves and is then relayed to us. Most of our information about the world is actually acquired this way – as second hand knowledge (Wilson, 1983). We experience firsthand very little of what we come to "know".

We depend on processing and sense-making done by others, in a different place, in a different time, to help us make sense of our world (sometimes to a polarizing degree (Gilovich, 1987)). This outsourcing of sense-making is fueled by necessity. We do not have the time or energy to collect, process, synthesize, and employ all our own data in a modern world. There is a division of labor and with it a division of knowledge and expertise (who was the last person to know "everything"?). To function in a (modern?) society, we depend on others, both past and present, for help when fulfilling our information needs.

And with this dependence on others, both in person and via the documents and records others create, we must also be wary. We must keep a vigilant eye towards the legitimacy of the information being passed along. We must evaluate, critically, the source and the provenance of second-hand information. Savolainen writes that when evaluating others and what we think they know, "overall, cognitive authority was characterized as having six facets; trustworthiness, reliability, scholarliness, credibility, 'officialness' and authoritativeness; of these, trustworthiness was perceived as the primary facet" (Savolainen, 2007, 3).

And even with the successful vetting and application of second-hand information, or shortcuts, from others, we never have perfect information. We may collect more information and we may collect better information, but it is never all the information we need to make perfect decisions. We satisfice; we satisfy with what is sufficient (Simon, 1957). We use what information we have to make decisions that we deem to be good enough at the time. We often seek out more information before making a decision but we have, what Simon called, **bounded rationality**. We have imperfect information, limited attention and money, limited processing power and limited time, but we still need to make decisions.

Choo's Decision Behavior Model (Figure 2.1) shows us that contextualized decision making happens within organizations based on cognitive limits, information quality and availability, and the values of the organization (Choo, 1996, 332). These inputs are handled with bounded rationality and within the confines of performance concerns, and whether the decision is good enough, among other simplifications. This decision making behavior is rationally expected and observed.

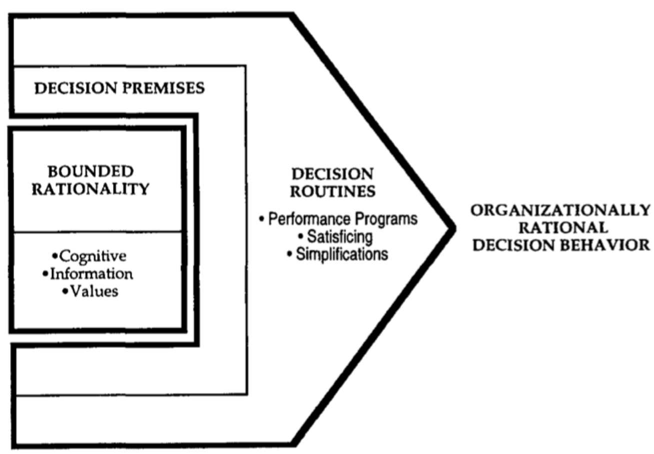

Figure 2.1: Choo's Knowing Organization, Decision Behavior Model

Even knowing we will never have perfect information when working in these limited environments, we can arguably make better decisions if we can improve or increase the amount of information on hand when making decisions. Having more good information reduces uncertainty about the environment surrounding a decision, but it does not necessarily reduce equivocality. To reduce equivocality, or ambiguity, of the information we have on hand, we need sensemaking and a perspective that comes from "retrospective interpretations" of earlier data and decisions (Choo, 1996, 334). We need to have *seen this before* and *know what it means*. What we need to make good decisions, in addition to good information, is called expertise.

There is a vast amount of latent, untapped information in the environment around us. Some of it is in the built world, some of it is in the natural world (too big, too small, hidden in non-visible wavelengths, etc.), and some of it is in the heads of those around us. If we are informed by the right people before making decisions, then we may improve our knowledge and understanding of a situation or problem at the time when we need to decide. **Knowing from whom we should get our information, when we are not sure of what we need, is a very hard problem.**

Expertise location, for this reason, has been a focus of the knowledge management field for many years. Knowledge management has also focused on the process of organizational learning and dissemination of that learning within the organization. In many cases, this has been done through the tracking of created documents and other knowledge artifacts (Martin, 2008).

An additional approach should consist of uncovering that which has not *yet* been recorded – that information which is in the heads of a group's membership. We should be equipped to hold up a mirror to help reflect an organization's insights and expertise back on itself. We need to help uncover the dark corners where we are not sure about the expertise in the room. With a regimen of self-reflection, iterated over time, I hope this problem can be made less hard. I think we can discover *who to ask* for the relatively low cost of a little sustained individual effort and some focused record-keeping in the distributed network.

2.3 Significance

When we are seeking answers to questions or trying to increase our knowledge in a certain domain, we seek sources of information that are credentialed and tested. We ask those who have come before us and who have learned from their own experiences – either through doing or through their own process of seeking and discovery. The sources we come to trust should have a history of providing good information in that domain in the past. We also come to expect them to continue to provide good information into the future. They should be known by others as keepers of good information and sound provenance. Our highly concentrated word for this set of qualities is reputation.

Those who have a good reputation perhaps spent many years developing their stature or physical skills in a field or domain. From the world of archival studies[1] [2], we know that physical and electronic sources of information should have a clear chain of custody and line of provenance as the document of record. If people are to be trusted as sources, as experts, we should be able to see the clear chain of custody and provenance of those who defer to these experts. Identifying these trusted human sources and the provenance to go with them is the thrust of this research.

Knowledge management has been about having the organization know what its members know. If this is synthesized a bit, we may talk of what the members *know about*. If we can reliably assume that a group can know what a person knows about, we can potentially do some very interesting things. We may be able to render moot the concerns we have today with individuals lying to increase their stature. If the group can reliably increase the social friction necessary to gain unmerited influence, we could safely ignore the opinions of those who have not convinced quite a few of his peers that he knows what he is talking

[1] Society of American Archivists' *Archives, Personal Papers, and Manuscripts*
[2] Canadian Council of Archives' *Rules for Archival Description*

about. Until he has as least the loose credential of a few peers who vouch for his credibility, his potential for abusing that credibility is severely limited. Of course, existing credentials, more formal credentials (diplomas, certifications, licensing, etc.), already allow this kind of credibility abuse. The addition of a loose socially awarded credential to the existing landscape would not affect the potential for abuse of those existing formal credentials. One would assume they would continue to convey more credibility than that provided by social labeling alone.

If a group can know what areas of expertise a person has, it may be able to better distribute articles for peer review to those who can best ascertain the quality of a pending publication. Important questions that arise could be distributed more reliably to those who could provide an informed opinion. Reporters in remote locations may be better able to determine who has actually been on the ground in Tehran during the presidential elections and who has recently created a Twitter account only to influence the placement of news articles during the next news cycle.

In a more formalized decision making process, voting systems could have weighted votes. If the matter at hand should not be decided democratically (e.g. one person, one vote), the relative weight of the votes could be set to match the relative weight of a voter's apparent relative expertise on the matter. This could mirror the practice of corporate elections based on shareholder totals. Those who know, instead of those who own, would be rewarded with influence. Perhaps just as interestingly, those who do not know could be ignored at vote tallying time. Internet-scale applications are often fraught with noisy comments and hostility. These could be programatically tuned out or weighted less if it was deemed useful or helpful to do so. And this could be done site-wide or customized for each viewer based on personal taste. It is also important to note that this type of filtering would be done post-hoc. It would not affect who could initially vote or comment. It would only affect how the display of the event would be rendered later. The original "democratic" vote totals would still be tallied and available.

But all of these scenarios depend on the assumption that a group's opinion about a member's areas of expertise can be trusted as "correct" – as good enough. The group's visible, shared opinion should allow the members of the group to make better, more informed decisions with less effort in less time.

I want to provide a robust means for allowing a group to assess and believe in their collective opinion about an individual's areas of expertise. They would be able to transparently evaluate how they grant cognitive authority to an individual and continually reflect on it. It would become a market indicator of what people know – one that fits into a larger, existing ecosystem.

This social reflecting lens should provide a form of **loose credentialing** and help to bring the implicit to the surface and make it explicit. When they choose to provide it, the trusted, focused, tacit knowledge in the heads of those we know could be available to all of us.

> "The grand challenge is to boost the collective IQ of organizations and of society" - Doug Engelbart regarding the Bootstrap Principle, a human-machine system for harvesting collected knowledge and evolving the technology for collective learning (Engelbart, 2004)

We still have far to go before the online and offline worlds truly merge. Eventually, we will enjoy a global transparent layer of data that is collectively curated and managed, but until that time, we continue to interact with other humans face-to-face much more often and in much more significant capacities. Lowenstein says that people trust their offline counterparts more than online social media (Lowenstein, 2009). However, research in computer-mediated communication (CMC) says we react to machines as people, at least subconsciously (Reeves and Nass, 1996), but we still have deference towards “real people” when we take the time to think through the communication event more carefully. When interacting with others via mediated channels, we usually do not focus on the medium itself and therefore we confer trust more than when the medium is explicitly obvious to us. As the media becomes transparent and easy and common, it will become more trusted.

2.4 Related Work

2.4.1 Expertise Location

Organizational Memory (OM) is a key component of Knowledge Management (KM). Abecker (Abecker et al., 1997, 1) writes “that an OM [system] has to be more than an information system but must help to transform information into action.” One part of OM is Expertise Location and Management (ELM), or the tracking of know-how within an organization (Lamont, 2003). As keeping track of employees’ knowledge is generally a very expensive undertaking for any size organization, a cheaper, more efficient technique for uncovering, managing, and disseminating this type of information would be a key contribution.

KM exercises involving human time and effort are naturally expensive for the firm. As such, incentivizing participation is one of the greatest hurdles to the implementation of a KM system (Ehrlich, 2003). Engaging with professional communities of practice (Lave and Wenger, 1991; Duguid, 2005), physical workspace reconfiguration, and encouraging water-cooler discussions can improve the sharing and awareness of expertise among professionals. Even so, Ling (Ling et al., 2009, 135) suggests that the single best type of incentives for knowledge sharing activities remain top-down such as “rewards and performance appraisal". Callahan agrees and suggests that managers must be involved, resources (time and money) must be given to the task, and overt (already known) content must be used to seed any initial system that hopes to elicit tacit content (Callahan, 2006b).

Stein (Stein, 1995) provides a standard set of stages for the understanding of organizational memory - knowledge acquisition, retention, maintenance, and retrieval. This is similar, but not identical, to Dieng's model for corporate memory management - detection of needs, knowledge construction, distribution, use, evaluation, and evolution (Dieng et al., 1999). Each suggests a time-lined progression but differ in that Stein's stages feel more institutionalized and less a collaborative effort. Dieng's use, evaluation, and evolution incorporate the dynamic nature and multi-person aspects of a distributed know-how.

Dieng (Dieng et al., 1999, 578) writes:

> However, the goal of a corporate memory building is different from the goal of an expert system: instead of aiming at an automatic solution for a task (with automatic reasoning capabilities), a corporate memory rather needs to be an assistant to the user, supplying him/her with relevant corporate information but leaving him/her the responsibility of a contextual interpretation and evaluation of this information (Kühn and Abecker, 1997). Kühn and Abecker (1997) notices that 'in contrast to expert systems, the goal of a corporate memory is not the support of a particular task, but the better exploitation of the essential corporate resource: knowledge' and cites some knowledge-based corporate memories (e.g. KONUS system aimed at support to crankshaft design).

Existing tools around Expertise Location and Management involve, almost entirely, self-description or existing-document data-mining (Lamont, 2003; Fitzpatrick, 2001; Becks et al., 2004; Balog et al., 2009). Traditional *tf–idf*[3] and bag-of-words analysis on these document stores can uncover a vast amount, but I think these techniques are missing out on what is in the heads of those who work with the person of interest. This is an important enough distinction to be made in a controlled environment, where the identities of the people involved are fairly well known and stable. However, trusting self-description in an unstructured, internet-wide environment without corporate identity management software seems ripe for abuse. The individual in question could easily be misrepresenting himself with malicious intent. Convincing many others of a lie or getting others to lie in a consistent manner regarding one's areas of expertise is much harder than deciding to lie on one's own behalf.

A tool for assisting in Expertise Location should meet the following requirements as set forth by Abecker (Abecker et al., 1997):

- gather information from multiple sources
- integrate with existing infrastructure and practices
- require little overhead in time/attention and provide benefits quickly
- actively present relevant information
- must stay up-to-date

Contextual Authority Tagging would handle the first, third, and fifth natively. Integration and presentation would both depend on implementation details. Ehrlich goes on to say that these systems must be fast, easy to use, engender trust in their results (e.g. be accurate enough to warrant continued use), and scale to the whole enterprise. Additionally, they must be used by management if the culture of the organization is expected to embrace the adoption of such a system (Ehrlich, 2003).

[3]Term Frequency–Inverse Document Frequency

2.4.2 Existing Systems

Systematically identifying the experts has been an ongoing research problem for quite some time (Ackerman and Malone, 1990; McDonald and Ackerman, 1998; Lutters et al., 2000; McDonald, 2001).

First, we have asked the people themselves to describe their own talents and areas of expertise, but this has demonstrated problems of motivation and incentive, as well as issues involving truthfulness and bias (Fitzpatrick, 1999; Yamim, 1996). Additionally, self evaluation leads to blind spots and the tricky pre-coordination problem of not knowing who the audience will be. We explain what we do and what we know differently to a colleague in the same field than to someone who does not already have a working knowledge of our own area. We contextualize when describing our skills to others face-to-face, because we can, because we know the audience. When asked to this for all possible audiences, we stumble.

Alternatively, we have investigated and analyzed the knowledge artifacts that have already been produced (Balog et al., 2009). Trying to identify the latent expertise from the documents that are produced and the transactions that have been recorded has been well studied, e.g., reports and meeting minutes (Craswell et al., 2001; Balog et al., 2006; Balog and de Rijke, 2008), email (Campbell et al., 2003), and social network analysis (Zhang et al., 2007; Balog and de Rijke, 2007). This area is also changing rapidly as we move into social spaces with our technology at increasing rates (corporate installs of social websites like Facebook, delicious, LinkedIn, Twitter). We are producing more artifacts than ever before, which is actually creating a different problem – there is too much. Finding the wheat is proving increasingly difficult.

Some existing systems include technology that allowed for both self documentation as well as automatic extraction and creation of profiles. The Community of Science's Expertise product allows for scientists in all fields to maintain an expertise profile that can follow them throughout their career, but the fields are self-updated and badly out of date or sparsely populated for many who have profiles in the system (Fitzpatrick, 1999, 2001). HP's internal Connex directory of experts also allowed for self-description and self-updating (Davenport, 1997; Becerra-Fernandez, 2000). The National Security Agency has an internal staffing and project matching system named the Knowledge and Skills Management System (KSMS) but is based on a custom knowledge taxonomy. Booz Allen Hamilton runs an internal expert skills directory that helps consultants match their expertise with clients' needs (Becerra-Fernandez, 2000). In 2008, Tacit.com sold their expertise location technology, based on automatic profiling from corporate email, and rolled their solution, illumio.com, into Oracle's Beehive collaboration platform. Cameron Marlow's Tagsona is Yahoo's unofficial internal directory that implemented tags and allowed employees to label each other. IBM built Fringe Contacts around the idea that people-tagging is a viable way to categorize "people's skills, roles, and projects in the form of a 'tag cloud'" and was modeled off the earlier IBM work on Dogear, a document tagging system (Farrell and Lau, 2006). Perhaps the most famous of corporate directories, IBM's BluePages house both company controlled information (lines of direct report, past and current projects, contact information) and persona information (controlled/populated by the employee him/herself) (Callahan, 2006a). Most recently Google acquired Aardvark (vark.com) and its question and answer routing technology that is based on semi-automatic expertise profile creation. Each of these systems, with the exception of IBM's Fringe Contacts, does not allow social labeling. They include only self-reported metadata or automatically generated metadata.

I propose another method.

I ask, can we not have people talk about what each other know, and create a new, social, shared knowledge artifact? It should not be directly derived from either the documents produced or by the person being evaluated; it should come from the people around the person of interest. It should come from tacit, social knowledge.

Can we create a knowledge artifact similar to the existing knowledge artifacts, but with a greater ability to encapsulate the here and now and to bend with time? Humans can synthesize a vast amount of context and provide better descriptors and categorize each other in more nuanced ways than perhaps any text mining or latent semantic indexing algorithm can. Even if it is not better, it may provide a different, important perspective not currently harvestable through automated means.

A socially created, shared artifact might quickly adapt to new terminology, new clusters, and see patterns that other systems might take longer to "see". It could be a *new* artifact, one that portends to be the current culmination of knowledge and synthesis. It could be a cutting edge reflection on the knowledge and expertise of a group in the moment.

I want to ask, and then enable, people to help create this new artifact.

2.5 Contextual Authority Tagging

Once we have a social reflecting lens to help us see what a person knows about, it serves as a jumping off point for powerful assessments and assertions. A validated socially robust system of categorized areas of expertise could be the foundation on which to build business and social services.

Can we imagine an ever-available data overlay of expertise? It could be the collective back wall that all ideas get bounced off of before further discussion – a back-chatter that has the opinions you value and need at any time. If it is ever-present and ever-evolving, it could influence nearly every decision we make when we interact with others. It could become the input we need to feel confident. We could eventually feel exposed and vulnerable without it.

It is important to remember that, as we move forward, we do not lose the ability to continue mining all our existing artifacts, documents, and logfiles. These are the raw materials that we use when we generate and manufacture our opinions. The socially constructed representation of one's areas of expertise, the visible version of Wegner's Transactive Memory (Wegner, 1986), would be a new source of information and would only serve to complement what we have already been able to do within the realm of document management (Choo, 1996). Keeping the focus on the people instead of the artifacts they create may better reflect the organizational knowledge inside a group and could greatly reduce the periods of time when new entrants are trying to get their bearings in a new office or managers are trying to assign relevant people to the task at hand.

Contextual Authority Tagging is a proposed technique for expertise location within a group by creating explicit knowledge from the group's individual tacit knowledge about each members' areas of expertise (Nonaka, 1991). This group can be an organization of any size, a loose affiliation of acquaintances or colleagues, or potentially everyone on Earth. For the purposes of this research, the scope of Contextual Authority Tagging will be directed towards the small and medium-sized working organization and membership. If and when this technique is shown as viable, then a greater scope could be approached, but at this time, some basic assumptions need to be questioned and verified.

Individuals have diverse interests, experiences, and connections with others. Some individuals have a wide variety of areas of expertise with working knowledge across many domains. Other individuals may live a very focused life and have extensive depth of knowledge in one area or two. As sources of information, members of each of these categories of individual are valuable, but in different ways. The Jack-of-all-trades may have insight into how techniques or methods fit together across traditional domain boundaries whereas the deep expert may have encountered a specific subtlety of something that one is beginning to work on and consulting with that person could save one lots of otherwise wasted time and money.

Knowing which people know which things is key to efficiently leveraging a network of contacts. Routing one's questions, seeking inspiration, and the building of teams each benefit from efficient use of existing mappings of knowledge and areas of expertise. Historically, these types of activities have been hard to commodify or automate. Humans are very good at applying a heuristic for knowing what others know and this research aims to tap into that talent.

Contextual Authority Tagging seeks to create and maintain a mapping of the areas of expertise of a network of individuals. It will do this by having the individuals involved use free text keywords or tags to label each others' areas of expertise. It is explicit and transparent and designed to uncover "reader-generated metadata" rather than "author-generated metadata". Results are shared back into the group and made visible, and the process is repeated. The resulting product is a weighted list of words associated with each person's areas of expertise. Words are weighted more heavily when more people used those words to tag an individual. Over time, the list, or some subset of the list (e.g. only tags from the most recent 12-month period), would presumably bend and follow the shape of the individual's current interests and knowledge as perceived by the group. Each individual's weighted list would be a specific fingerprint in the multidimensional space created by all possible keywords and could potentially serve as inputs and be used by a multitude of other tools to aid or weight in further decision-making tasks.

CAT is contextual in that each person's fingerprint is unique and relative both to the querier's network and to the queried's network. Limiting whose "votes" count could preempt noisy or "spammy" results. Limiting "votes" with respect to the time they were recorded could prevent "old" or outdated results.

One could imagine future algorithms working in the background, being recursive in nature (similar to Google's PageRank (Brin and Page, 1998) or Kleinberg's HITS (Kleinberg, 1999)), returning a ranked list of people as weighted by how many other people, who have weight in that domain, "voted" for those listed.

Authority refers to the cognitive authority being granted by the network to each group member (Wilson, 1983). Wilson differentiated between administrative authority (which is obtained by virtue of position or rank) and cognitive authority (which is granted by others based on experience and demonstrated knowledge). The fact that this authority is granted, rather than held "ex officio", is what makes CAT interesting.

The opinions of one's peers hold interesting collective insights and this technique hopes to tap into this insight and bring it out where both the individual can benefit from her own hard work and expertise and others can more efficiently locate that expertise.

2.6 Research Questions

Contextual Authority Tagging has been conceived and designed to get at two major questions regarding how a group comes to know about its own areas of expertise. The following questions are raised and will be addressed by the proposed research methodology.

R1. Does it work?

a) **Similarity** - How similar are a group member's opinion of his/her own areas of expertise and the group's opinion of his/her areas of expertise?

b) **Convergence** - How does the similarity behave over time? Do the two opinions converge? If so, how long does it take? If not, is there a persistent gap?

R2. Does it matter?

a) **Comfort** - How comfortable are group members in participating? What are the main factors influencing their comfort level?

b) **Confidence** - How confident are group members in a system like this? Does this system provide a valid credential? Does this system increase users' trust in one another?

c) **Usefulness** - What is useful about a system like this? What did participants learn? Are participants satisfied? How would using this system affect participants' decision making?

Latour and Nelson suggest to us that where there is a lack of contention, a social fact will be defined (Latour and Woolgar, 1986; Nelson, 1993). Social tagging phenomena have demonstrated a stabilization of tagging behavior (Russell, 2006; Golder and Huberman, 2005). Together, these suggest the first hypothesis:

H1. As the social fact of what a person knows is molded by the group, a consensus will appear and converge.

The comfort levels of the participants will depend on their surroundings, the familiarity of the task, and their feelings of control:

H2. Comfort levels will increase as the system becomes known and understood. Initial trepidation will be assuaged as the system allows participants to see more of how they are perceived by others.

The warranting principle suggests that we give more credence to information provided by others, rather than information within the control of a particular other (Walther and Parks, 2002; Walther et al., 2009). Online or offline, information that is known to be easily manipulated is less trusted. Additionally, Delphi-style studies increase the confidence levels of the participants (Rowe et al., 2005). This leads to the third hypothesis:

H3. Group members will have confidence in this system and exhibit increased trust in one another.

Chapter 3

Proposed Study

3.1 Overview

I am interested in exploring the ability of a group to identify the areas of expertise of its members.

Current efforts to capture this type of information almost always derive their value from either documents produced by or between group members or from asking members to talk about themselves and their own areas of expertise and knowledge. The document method is at the basis of most expert systems and knowledge management software of the past couple decades.

The self-disclosure method works, at best, when all the members tell the truth, have the best interests of the group at heart, and are thorough in their descriptions of their skillsets and knowledge. Usually, data of this kind is simply too sparse or outdated to be actionable. Members may leave out important items from their descriptions or not participate at all. Worse, members may simply lie about their skillsets for any number of reasons.

A more robust system may be available by allowing the members to talk about each other. Holes (where things were left out) may be filled, and deception would be made more difficult because many in the group would need to give false descriptions for the collective opinion to be swayed.

If a group can (or does) know better than an individual, there should be a way to ask them. Contextual Authority Tagging may allow for the systematic gathering and evaluation of this type of information.

3.2 Delphi

The proposed study would use a modified version of the Delphi method. The original Delphi study was run in the 1950s and 1960s by the RAND corporation to help the US Government determine the nuclear capabilities of the Soviet Union (Helmer and Rescher, 1959; Dalkey and Helmer, 1963). They were studying the unknown military futures market by asking a variety of experts to answer a battery of questions. The answers were collated and then distributed back to the experts for additional rounds of answering the same questions - but critically, with the collective opinions of the other experts to aid their synthesis.

Rowe writes that, "in particular, the structure of the technique is intended to allow access to the positive attributes of interacting groups (knowledge from a variety of sources, creative synthesis, etc.), while pre-empting their negative aspects (attributable to social, personal and political conflicts, etc.)" (Rowe and Wright, 1999). Over the following four decades, the Delphi method has been refined and used in many other areas besides military futures, including social science predictions (Linstone and Turoff, 1975; Rowe et al., 2005; Hsu and Sandford, 2007).

Most research has suggested that with proper preparation and consideration for expert subjects, questionnaires, and evaluation, a Delphi study can run from three to five rounds, with four being the most common number of iterations (Hsu and Sandford, 2007). Some prior Delphi studies have used post-task surveys to sample participants' reactions - from satisfaction (Van De Van and Delbecq, 1974) to confidence (Scheibe et al., 1975; Boje and Murnighan, 1982) to difficulty and enjoyableness (Rohrbaugh, 1979) - and I plan to employ some of the same types of questions with CAT, especially considering the subjects are being asked to formalize their informal knowledge about one another.

A traditional Delphi study involves 1) an objective facilitator who gives "controlled feedback" in the aggregate, 2) a collection of independent experts in a domain (anonymous, to each other), and 3) a series of evaluations (iterations) designed to have the collective opinion of the experts predict the future in that particular domain (Rowe and Wright, 1999).

I want to modify this method, to instead, have members of a group or team define the areas of expertise for one other. This substitutes for the original formula 1) a piece of software to facilitate and aggregate free-text tags from 2) the members of the group who are anonymously tagging each other's areas of expertise in 3) a series of rounds where cumulative tagging information is visible from prior rounds. A group of ten members would be, effectively, running ten concurrent Delphis at one time – all of the participants evaluating each of the participants.

Documented criticism of the Delphi consists of lack of statistical tests, lack of demographic description of the participants, the eligibility and selection of the expert participants, the lack of explanatory quality of the responses, and the degree of anonymity of the participants (Luo and Wildemuth, 2009).

Additionally, Delphi studies need to be carefully administered to avoid the following things (Linstone and Turoff, 1975):

- overspecification of the problem statement and potential dampening of diverse perspectives
- inadequate summarization during the aggregation and synthesis stages
- lack of common interpretation by the participants of any scales being

applied

- ignoring of differences of responses among participants that could be fruitful
- underestimating the amount of time and effort required to participate and administer the study
- misunderstandings between participants due to cultural or linguistic differences

Delphi has a lot to offer as a grounded, tested method to find convergence of opinion given its skeleton of domain experts, anonymity, and iteration. As Contextual Authority Tagging is being proposed to help uncover (unleash?) a collective subjective truth, the Delphi method seems appropriate as a construct upon which to formalize the proposed research.

3.3 Modified Delphi

I intend to ask a team of people about their opinions, aggregate their opinions, redistribute their opinions back to the group, and then iterate the process. This process should continue either for a minimum amount of time, until their opinions "converge", or until a maximum amount of time or iterations has been met.

As opposed to a traditional Delphi Method study, wherein the participants are selected and recognized as experts and the point of the study is to identify their collective opinion on a matter, this study will be using groups of people who work with one other. These group members, while not necessarily experts in any specific domain, know each other well enough to describe each others' areas of knowledge and expertise. They already grant some cognitive authority to each other in certain areas, and this study will ask them to explicitly name those areas.

The irony of looking at expertise with a method originally designed to use experts is not lost (or intentional), but I do think the thinking holds up. An individual's colleagues spend more time thinking about what that individual knows more than probably anybody else, outside of the individual herself. They are uniquely situated to evaluate the question around the individual's areas of expertise – and therefore, I'm considering them the equivalent of the selected Delphi experts. Traditionally, this type of expertise evaluation has been done either solely by the individual (via her résumé) or her boss (in a letter of recommendation or reference). I hope to add a potentially useful voice to this duo.

Through the anonymous aggregation and redistribution of the group members' descriptions, the areas of cognitive authority will be named and quantified by the group. I will be employing simple keyword labeling, or tagging, as the method by which group members will attribute areas of expertise to one another.

Some limitations and concerns are addressed in the following subsections.

3.3.1 Anonymity

Any concern over the **anonymity** of the participants or the attribution of the tags is reduced to the security issues around the database where the information will be stored. For research purposes, I plan to store both the "tagger" and the "taggee", but this would not be strictly necessary if plausible deniability was of due import. Further concerns over who said what are relegated to the realm of the social – the scope of which is beyond the aim of this research. I am assuming that by making these expertise tags visible and available for discussion, some stories regarding the provenance and justification of the tags will be told. Truly secret information should remain secret, regardless of the availability of a tool or exercise like CAT – but that is an issue between those who have secrets (or privileged/private information) and those who know the secrets.

3.3.2 Selection

I am also expecting to hear feedback regarding the **selection** of participants of the form "friends/colleagues are not experts." I posit that they are expert, within the context in which the experiment is run. Some information about a participants' areas of expertise will surely be beyond the purview of the other participants involved, regardless of the environment in which the experiment is run. That said, I am limiting the research to be run in professional or collegial environments where intellectual activity is the main type of interaction between participants. I am explicitly avoiding, at this time, groups that could be construed as family, social, hobbyist, or athletic. By sticking to offices and workplaces, I expect that the types of information generated by CAT to remain largely "on topic" as that is the nature of the majority of interactions between the participants. Additional "off topic" information would be generated and displayed as well, but it is expected that these would be limited in scope and not extend much beyond what is commonly discussed at work already; the participants will continue working together after the experiment is complete. And of course, with further consideration, any "off topic" information could be removed in later rounds.

3.3.3 Misinformation

There will probably be concern over the possibility of **negative information** or **false claims**. These two concerns are important and deserve attention. I suspect that non-normative behavior and aberrant tags will draw attention quickly. This is no different from unprofessional language being uttered or a physical disruption in the workplace – it is quickly noticed and addressed. Negative tags will largely be disincentivized by the positive phrasing of the question being asked, “What do you think this person knows about?”, and “What are this person’s areas of expertise?”. Answers like “being a jerk” would stand out and not be corroborated by others over time. That said, if it *was* corroborated and voted up by others, this tag arguably is doing a service to the community by making it apparent to this participant that they are not viewed as helpful by a contingent of their peers and co-workers. This could, arguably, lead to better behavior on the part of the tagged.

3.3.4 Coverage

Another concern that could be leveled regards coverage of the generated tags, and the fact that there would remain **hidden information** not captured by this technique. I agree, but do not see that as a limiting factor. I assume that humans will always hold some information to themselves – and I encourage that. I also think that the anonymity provided by CAT will allow more information than is currently being put on display to be captured and propagated around. I think having total information would be a horrible thing. I also think that having a place for anonymous speech is important and that it sometimes brings potentially fascinating and useful information to the fore.

3.3.5 Statistical Rigor

Regarding the **lack of statistical measurements** and tests to determine the significance of the findings rendered by classical Delphi, I feel CAT can be claimed as immune. The nature of Delphi is that it results in a set of findings or opinions that have been deemed “convergent.” The weakness of these findings can be attacked from a predictive standpoint, but as I intend for CAT to be run continuously (if implemented beyond my dissertation research), the notion that a test was not conclusive or that there is no test is a non-issue as there are never any final “findings.” The co-workers will take what they want from the information and use it accordingly. I see CAT being a piece of reporting/learning infrastructure that allows other tools to be built and used around it for decision making. Making the opinions of people visible should create more opportunities for discussion and reduce the chance for misunderstandings.

3.3.6 Loss of Control

I also expect to see some pushback from (potential) participants regarding their **not having a say** in what is being said about them and the fact that this information is being published for others to see. My counterpoint is that this is already happening, everyday, all around us. People gossip and talk amongst themselves. CAT will just bring this information together, aggregate it, and show it publicly. Damaging gossip is gossip that happens anonymously and behind closed doors. CAT is done in the open. Those who are good at what they do, and know their stuff, will be rewarded. Those who have not convinced their colleagues of their areas of expertise will have sparse data to show for it. Additionally, those who are well liked will probably be rewarded more than those who are not. This is not as much a privacy concern as it is an issue of control. CAT, I agree, definitely moves the control of defining ones areas of expertise away from the individual and towards the group (but it does not remove the voice of the individual, it just adds the voice of the group). But I also think that moving control towards the group is a good thing and something we need as we begin to live in an ever-connected, online environment where notions of identity are not as ingrained and well-understood as in our known physical world.

3.4 Lists of Tags

Conceptually, CAT employs two sets of lists - created and processed. Created lists are lists that are created by members of the group, and processed lists are lists that are the output of the process of the exercise. They contain the same type of content (tags), but are shuffled, ordered, and aggregated differently as part of the exercise.

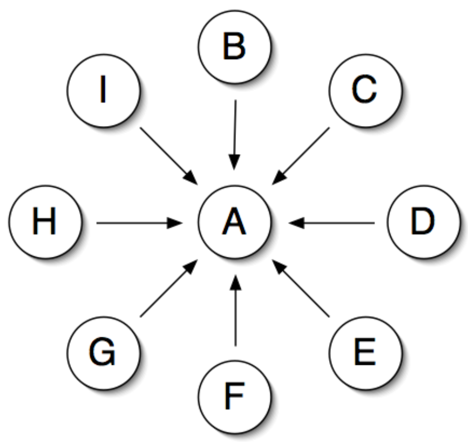

Figure 3.1: Canonical Group: Group member A is tagged by the other individuals in the group ($B..I$) and represented as $A_B..A_I$. Collectively, the tags generated by this group would be represented as A^*.

The first set are the lists as they are created by a member of the group of size n. For each iteration of the exercise, each group member creates n lists. They are of two types (totaling n): Self (1) and Other ($n-1$).

1. **Self** (A_A) : a list consisting of tags that a member uses to describe his/her own areas of expertise. There is only one "self" list, per member, per iteration.

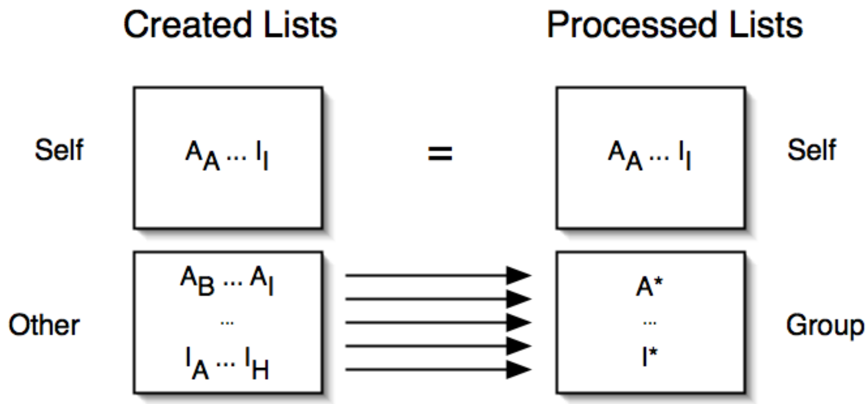

Figure 3.2: Four Lists: A group member creates lists about him/herself (A_A) and other group members ($B_A..I_A$). After processing, each member has a list about themselves (A_A) and what the group thinks he/she knows about (A^*).

2. **Other** ($B_A..I_A$) : a set of lists created by the member to describe each of the other $n-1$ members of the group. If there are 9 total members of the group, there are 8 "other" lists created, per member, per iteration.

The second set is just a reorganization by the system of the created set of lists. This set consists of the lists "about" a member, rather than "created by" a member. Each member of a group will have two processed lists describing them, for each iteration of the study.

The two processed lists include:

1. **Self** (A_A) : a list describing the individual by the individual (identical to the created "self" list above), and

2. **Group** (A^*) : a weighted aggregated list where the other group members describe the individual (the combined "other" lists into one)

If a group has 9 members (as in Figure 3.1), the first iteration of the exercise will generate a total of 18 processed lists, 2 for each person. If there are 5 iterations in the exercise, a total of 90 processed lists will be generated.

Within each iteration, or round, a series of four steps will be followed by each member of the group. The steps include:

1. **Review** : The member will be presented with the current state of the experiment from his/her perspective. His/her accumulated Self list and Group list will be visible. Self and Group lists will also be visible for every other member of the group. This is where most of the learning and consideration of new information presented by the tagging instrument will take place.

2. **Self Assessment** : The member will add and remove tags to the current "Self" list of tags.

3. **Group Assessment** : The member will add and remove "Other" tags for each of the other members of the group.

4. **Round Complete** : The member will be notified of completion of the current round.

These steps would directly follow one after the other in one sitting. The spacing of the rounds will be up to the groups themselves and could range from a few minutes to a few days. I expect most groups could finish five rounds within a two week window.

The examples in Figures 3.3 through 3.11 are shown as occurring in the fourth round of iterations.

3.5 Recruitment

I am planning to recruit 5-10 professional groups consisting of 8-10 people each. I plan to recruit these groups of co-workers via personal connections and existing offers to help (I have been discussing this idea for a couple years with professional contacts and have 5-6 outstanding offers to make contact when the study is ready to proceed).

I have performed a pre-IRB pilot test-run of the tagging software with both my circle of close friends as well as fellow SILS PhD students. The data from this pilot has informed my current design and is represented in the screenshots seen earlier.

A population consisting of the groups listed above should allow me to make sufficient claims about the nature of small working groups. I will seek a diversity of domains from these 5-10 professional groups and hope that sufficient breadth will satisfy those who may question the validity of such a small sample of groups. Each individual can be considered the subject of a Delphi study, and therefore, with 50 participants, I would have 50 data points.

I am not planning to compensate the participants of this research at this time. The participants and the companies they work for will be provided with the output of their sessions with me and will hopefully find the results insightful as is.

3.6 Instruments and Datasets

This research will be carried out in three major stages. Data in Stage 1 will be collected via the custom tagging software and an integrated survey. Data in Stage 2 will be collected through a set of semi-structured interviews. Data in Stage 3 will be comprised of both human- and algorithmically-generated similarity scores.

3.6.1 Stage 1a : CAT Software (Lists)

The custom software (seen below) will generate the primary tagging dataset for this research. A group of 8-10 people will use this software through 5 rounds and populate a database of taggings (rows) consisting of a tagger, a tag, a taggee, and a timestamp. When the experiment is complete and the software has processed the tagging activity, it will generate 1 list of tags for each combination of group, participant, listtype (self/group), and round. For an experiment consisting of 6 groups with 8 participants each moving through 5 rounds, the system will generate a total of $6 * 8 * 2 * 5 = 480$ lists. 2 of the 480 lists can be seen in Figure 3.4, one for self and one for the group.

3.6.1.1 Screenshots

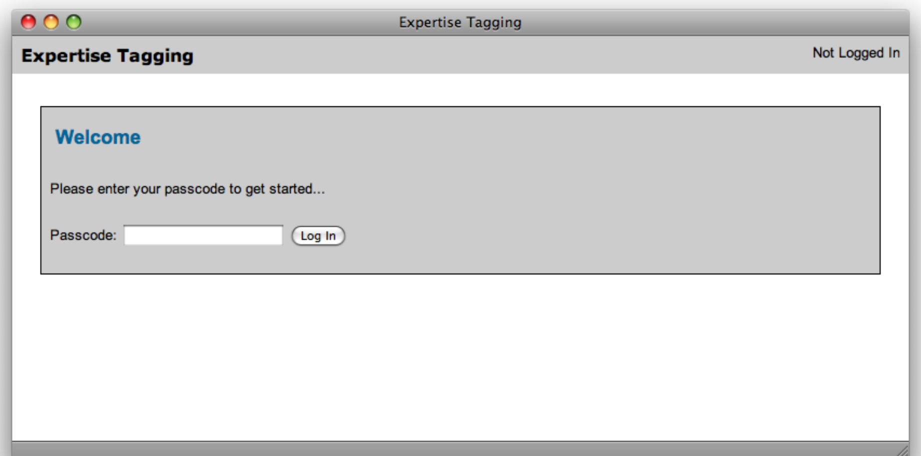

Figure 3.3: Prototype: Login: Each group member will use a simple passphrase to log into the system.

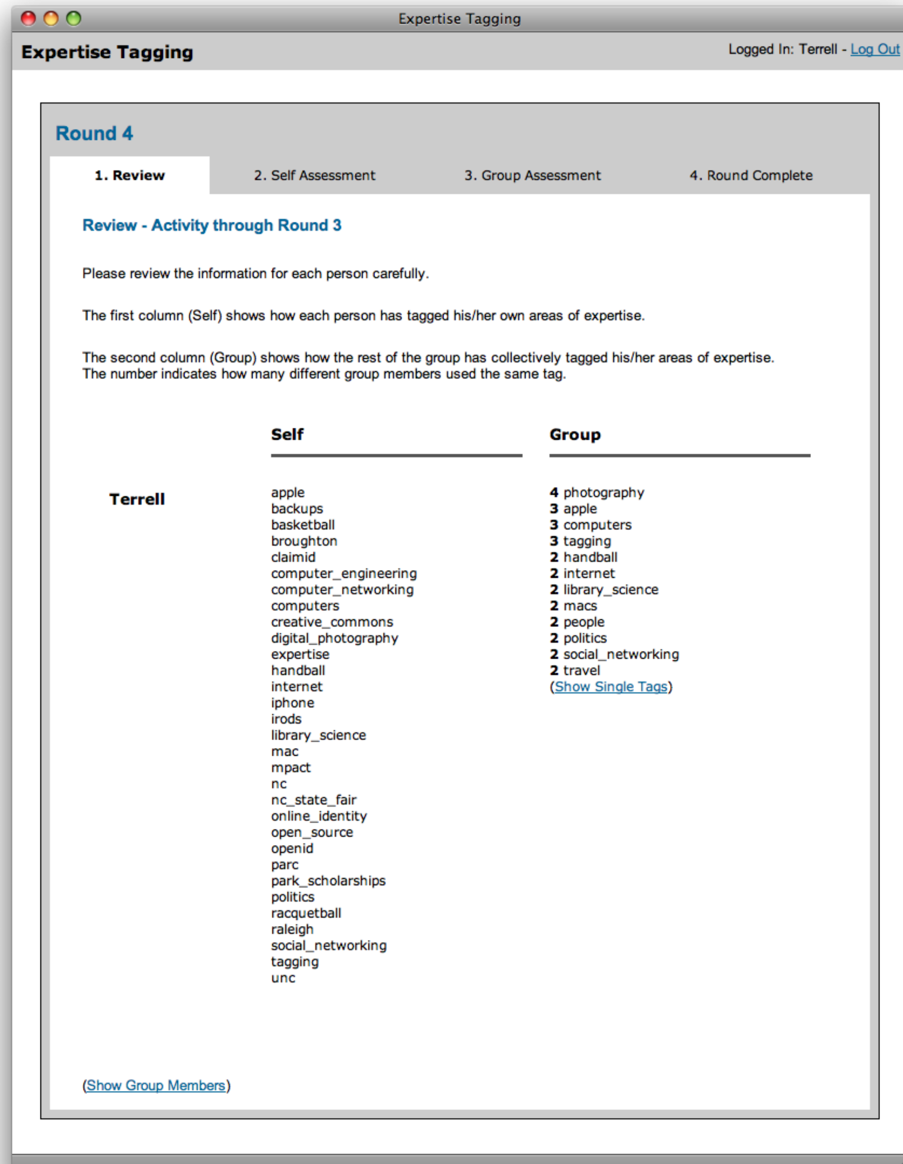

Figure 3.4: Prototype: Step 1 - Review - Self: The group member will be shown the aggregate listing of tags since Round 1. This includes both self tags and the aggregated tags that the group has put into the system about his/her areas of expertise.

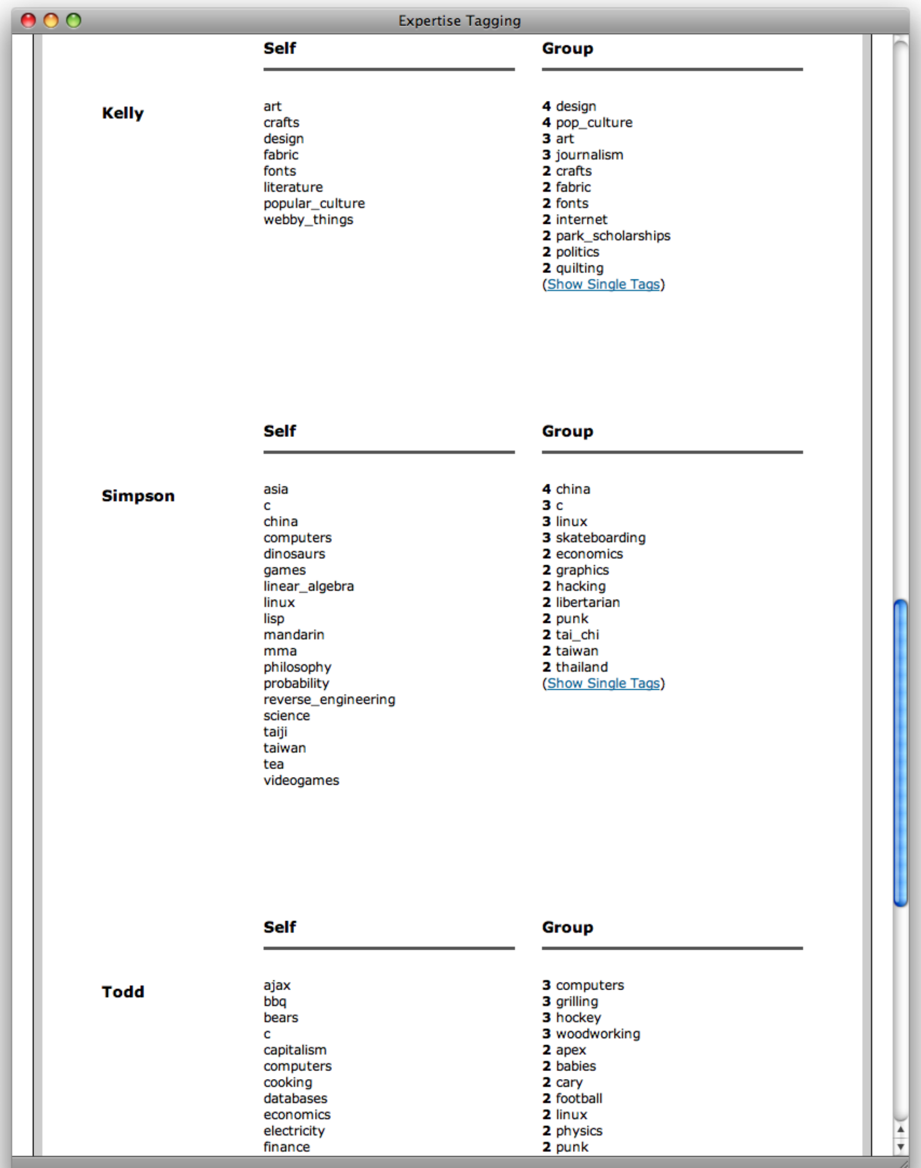

Figure 3.5: Prototype: Step 1 - Review - Others: The group member will be shown the aggregate listing of tags since Round 1 for each of the group members. These include both self tags and the aggregated tags that the group has put into the system about each group member's areas of expertise.

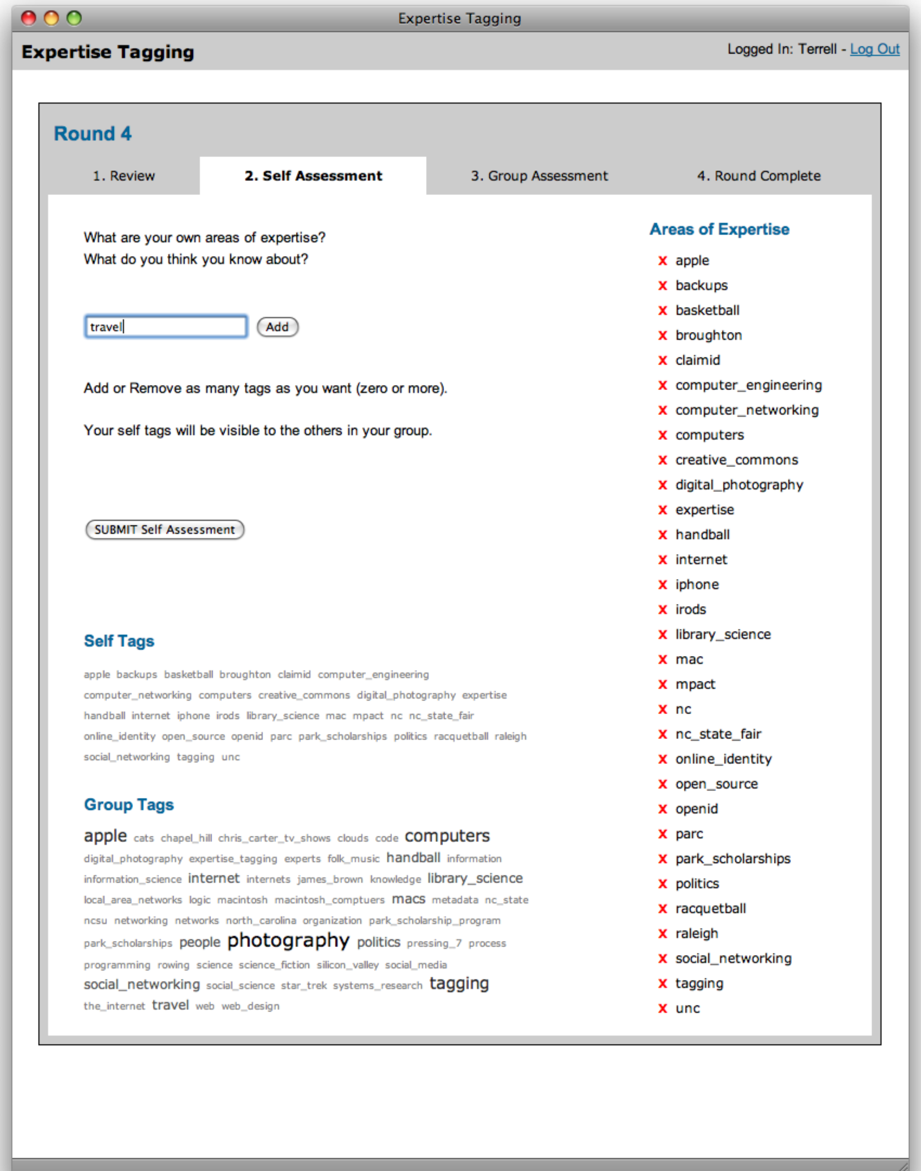

Figure 3.6: Prototype: Step 2 - Self Assessment: The group member is asked to tag his/her own areas of expertise. The full listing of existing self tags (from prior rounds) is shown in the right column. Any existing tag can be removed by clicking on the corresponding red X.

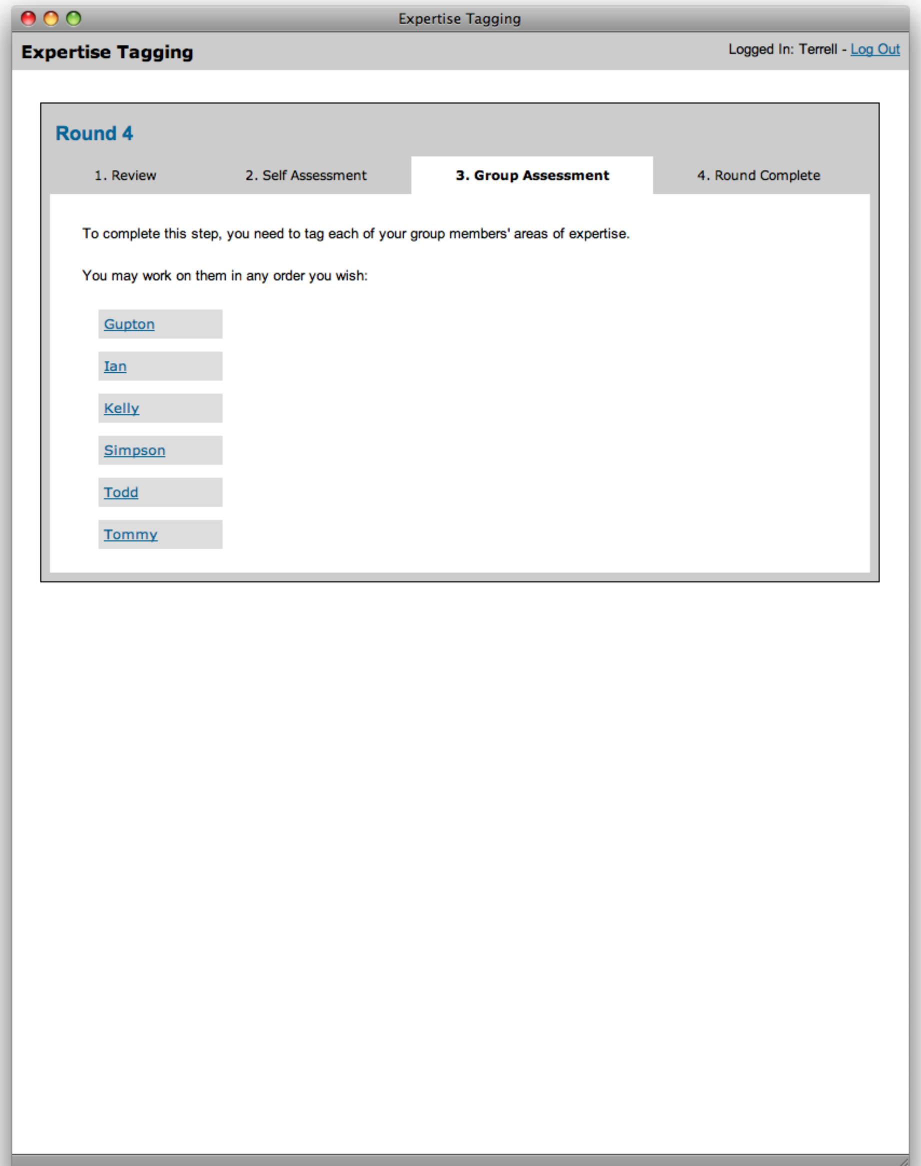

Figure 3.7: Prototype: Step 3 - Group Assessment - Before: The group member is asked to tag each of his/her group members during Step 3. The group members can be tagged in any order. The group members must all be "visited" before moving to Step 4.

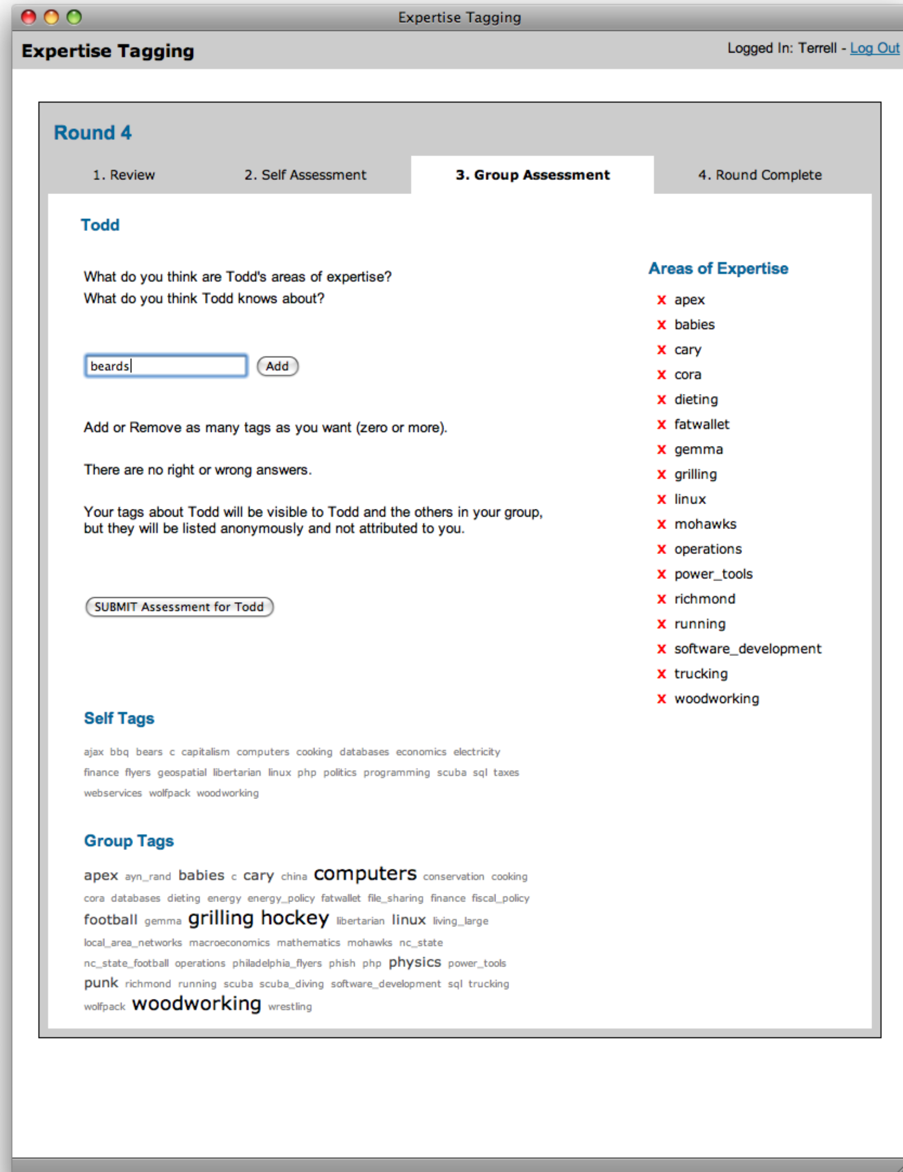

Figure 3.8: Prototype: Step 3 - Group Assessment - Tagging: The group member is asked to tag this group member's areas of expertise. The full listing of existing tags from the logged in group member (from prior rounds) is shown in the right column. Any existing tag can be removed by clicking on the corresponding red X.

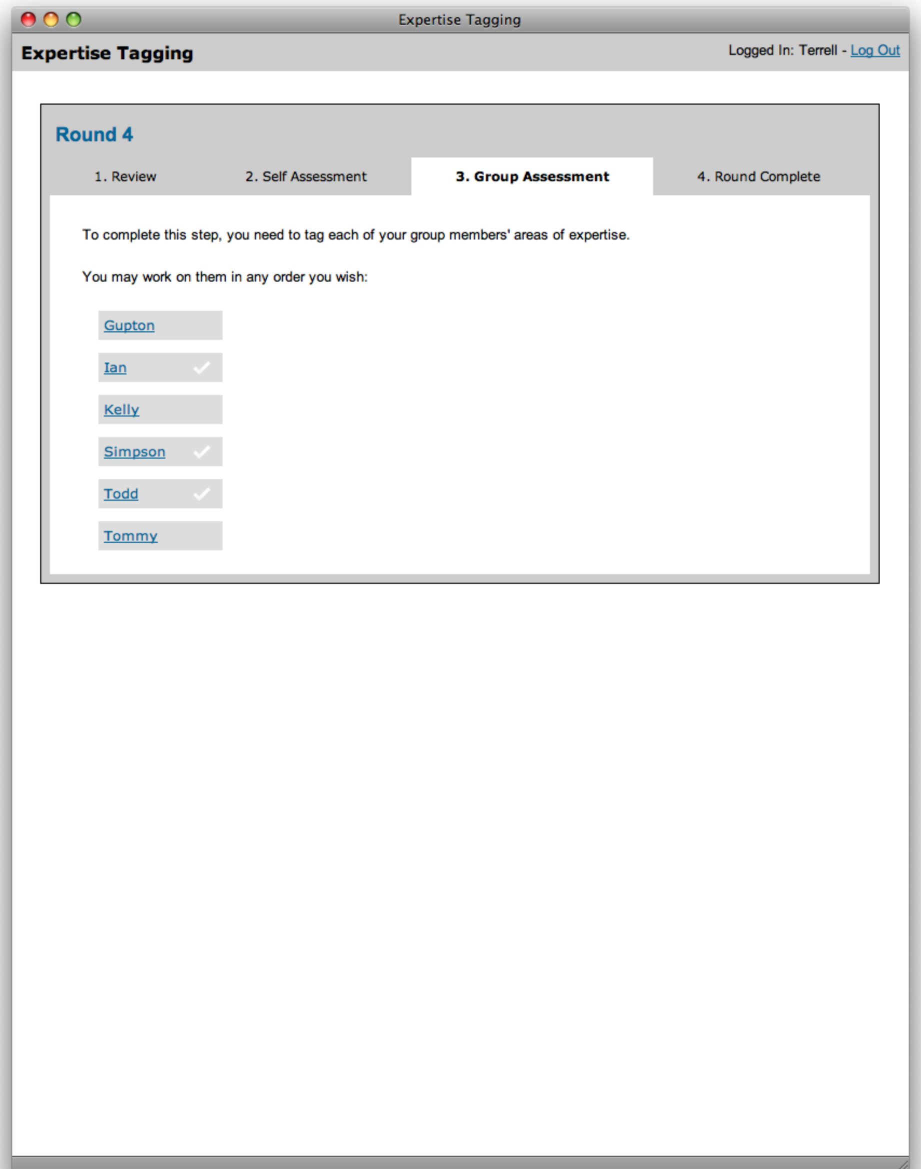

Figure 3.9: Prototype: Step 3 - Group Assessment - Partial: The logged in group member must “visit” each other group member before moving to Step 4. This user has tagged 3 of 6 of his fellow group members during this round.

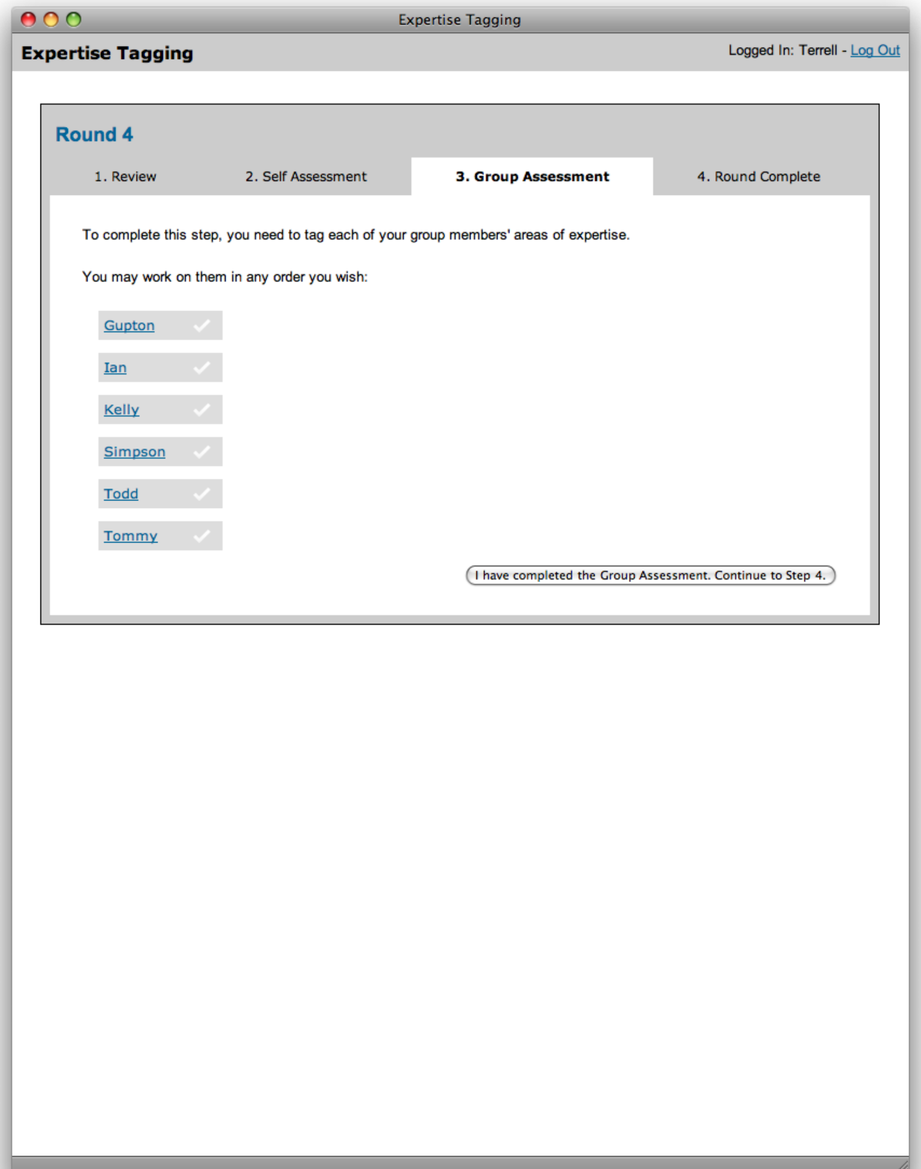

Figure 3.10: Prototype: Step 3 - Group Assessment - Complete: Each fellow member has been tagged in this round. The logged in member is ready to move to Step 4.

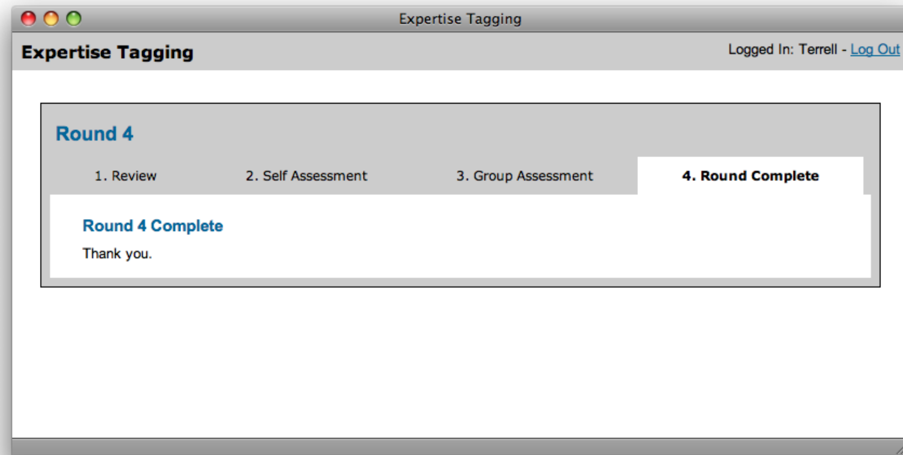

Figure 3.11: Prototype: Step 4 - Round Complete: The logged in member has completed this round.

3.6.2 Stage 1b : Survey

The survey that will be administered to each participant will be broken into a pre-test and a post-test. It is designed primarily to answer parts of the Comfort (R2a) and Confidence (R2b) research questions. The dataset generated by the survey questions will consist of 1 row per participant. Using the earlier example experiment, we would expect $6 * 8 = 48$ total responses.

3.6.2.1 Pre-Test

1. **How long have you been a part of this group?** (R2a) Less than 6 months, 6-12 months, 1-3 years, 3-5 years, More than 5 years

2. **How familiar are you with your group members' areas of expertise?** (R2a) Not Very Familiar, Somewhat Familiar, Familiar, Very Familiar, Extremely Familiar

3. **How familiar are your group members with your areas of expertise?** (R2a) Not Very Familiar, Somewhat Familiar, Familiar, Very Familiar, Extremely Familiar

4. **How trustworthy are your group members?** (R2b) Not Trustworthy, Somewhat Trustworthy, Trustworthy, Very Trustworthy, Extremely Trustworthy

3.6.2.2 Post-Test

1. **How familiar are you with your group members' areas of expertise?** (R2a) Not Very Familiar, Somewhat Familiar, Familiar, Very Familiar, Extremely Familiar

2. **How familiar are your group members with your areas of expertise?** (R2a) Not Very Familiar, Somewhat Familiar, Familiar, Very Familiar, Extremely Familiar

3. **How trustworthy are your group members?** (R2b) Not Trustworthy, Somewhat Trustworthy, Trustworthy, Very Trustworthy, Extremely Trustworthy

4. **How comfortable are you with your group's tags about your areas of expertise?** (R2a) Very Uncomfortable, Uncomfortable, –, Comfortable, Very Comfortable

5. **How comfortable would you have been if the system had not been anonymized?** (R2a) Very Uncomfortable, Uncomfortable, –, Comfortable, Very Comfortable

6. **How confident are you that this system gives you good information?** (R2b) Not Confident, Weakly Confident, Confident, Strongly Confident, Very Strongly Confident

7. **How confident are you that this system gives you new information?** (R2b, R2c) Not Confident, Weakly Confident, Confident, Strongly Confident, Very Strongly Confident

8. **How willing would you be to make decisions based on this system's output?** (R2b, R2c) Not Willing, Weakly Willing, Willing, Strongly Willing, Very Strongly Willing

9. **How useful was this exercise?** (R2c) Not Useful, Somewhat Useful, Useful, Very Useful, Extremely Useful

10. **How interesting was this exercise?** (R2c) Not Interesting, Somewhat Interesting, Interesting, Very Interesting, Extremely Interesting

11. **What was your favorite part of this exercise? Why?** (R2a) Free Response

12. **What was your least favorite part of this exercise? Why?** (R2a) Free Response

3.6.3 Stage 2: Interviews

The semi-structured interviews will be conducted with one member of each group at the end of the experiment (probably the liaison with which I have been in contact). The questions will be asked and responses will be recorded for later transcription and content analysis. Considering one interview per group, the example experiment would generate 6 interviews. These questions are designed to primarily answer the research questions around Usefulness (R2c).

1. What was your general impression of this exercise?

2. What did you learn about yourself? (R2c)

3. What do you feel the group learned about you? (R2c)

4. What was your favorite part of this exercise? Why? (R2a)
5. What was your least favorite part of this exercise? Why? (R2a)
6. How did the group feel about participating? Were they nervous? Excited? (R2a)
7. Was the exercise a success? Has it had any effect on how the participants act towards one another? (R2c)
8. How do you think the exercise would have been different if the tags had not been anonymous? (R2a)
9. Would you recommend this type of activity to others? To partner organizations or groups? Why or why not? (R2c)
10. Is there anything else you would like to share about this activity?

3.6.4 Stage 3: Similarity

The main thrust of this research is to determine whether a group and a particular member agree on a member's areas of expertise. The ratings from this dataset will be used to determine this level of agreement.

The similarity dataset is designed to describe the level of similarity between the different taggings lists that are generated by the CAT software. Evaluation of subjective information (such as one's areas of expertise) must be carried out in a relative manner - as there is no objective ground truth or known yardstick against which to measure.

Two separate methods of capturing this similarity will be used. The first will use humans to judge the similarity of the presented sets of words. The second will use an existing algorithm designed to find the semantic similarity between two sentences, but without using the information encoded in the sentence structure.

Graphing similarity scores against the iteration (round) will show whether the two lists (self and group) for a single person converge (become more similar) over time. If the similarity scores increase, then there is tendency towards convergence. If the scores do not increase, or they plateau, then there remains some difference in the lists and therefore, for a pairing of self/group lists, the self and group did not agree on that participants' areas of expertise.

This analysis can be performed for each person, then pooled and performed for each group as a whole, and then for the entire experiment. Analysis at each of these levels may prove interesting. If the individual graphs prove similar, the group pools will be representative. If the different group graphs prove similar, then the entire aggregate pool may prove generalizable to an even greater population.

3.6.4.1 Human Judged

This dataset will be generated via Amazon's Mechanical Turk. Mechanical Turk is a workplace where humans have enrolled to complete simple tasks for payment. It is designed to be used for tasks that computers are ill-equipped or too expensive to solve.

Two lists of tags will be shown to a Mechanical Turk worker and the worker will be asked to rate the two lists' similarity and the worker's confidence in their similarity rating (see Figure 3.12). The similarity and confidence ratings will be on five-point Likert scales. Each pair of presented lists will be evaluated ten times (by different workers) to check for consistency and reliability. These ten pairwise comparisons between "self" and "group" lists per participant per round would generate $10 * 6 * 8 * 5 = 2400$ similarity ratings (rows).

The two lists that will be shown will be the two lists about a particular participant from a particular round. The "self" list will be comprised of all the self tags from that participant, randomized and shown in no particular order. The "group" list will be comprised of all the common tags (used at least twice by the group), unweighted, and randomized. The randomization and unweighting is preferred to any other combination of presentation option as there is no way to otherwise represent the "self" and "group" lists equally. Tags only used once by the group are removed as they do not represent any type of consensus within the group and would greatly increase the size of the list to be displayed. (Alternately, I could run an additional round with the full group listing, or even a second additional round with weightings visible).

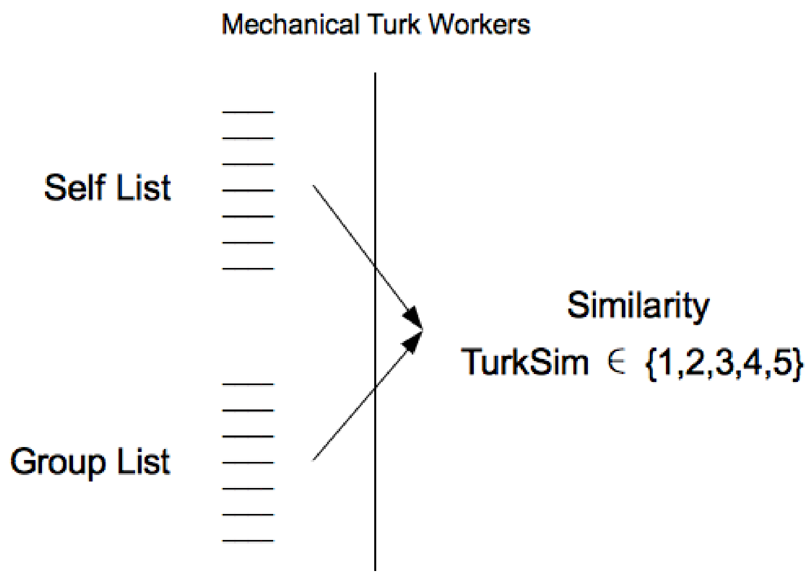

Figure 3.12: Human Similarity Rating Model – Two lists of words are presented to Amazon Mechanical Turk workers who are asked to rate their similarity on a 5-point Likert scale.

The following (Figure 3.13) is a simple rendering of the Mechanical Turk HIT (Human Intelligence Task) that will be presented to workers.

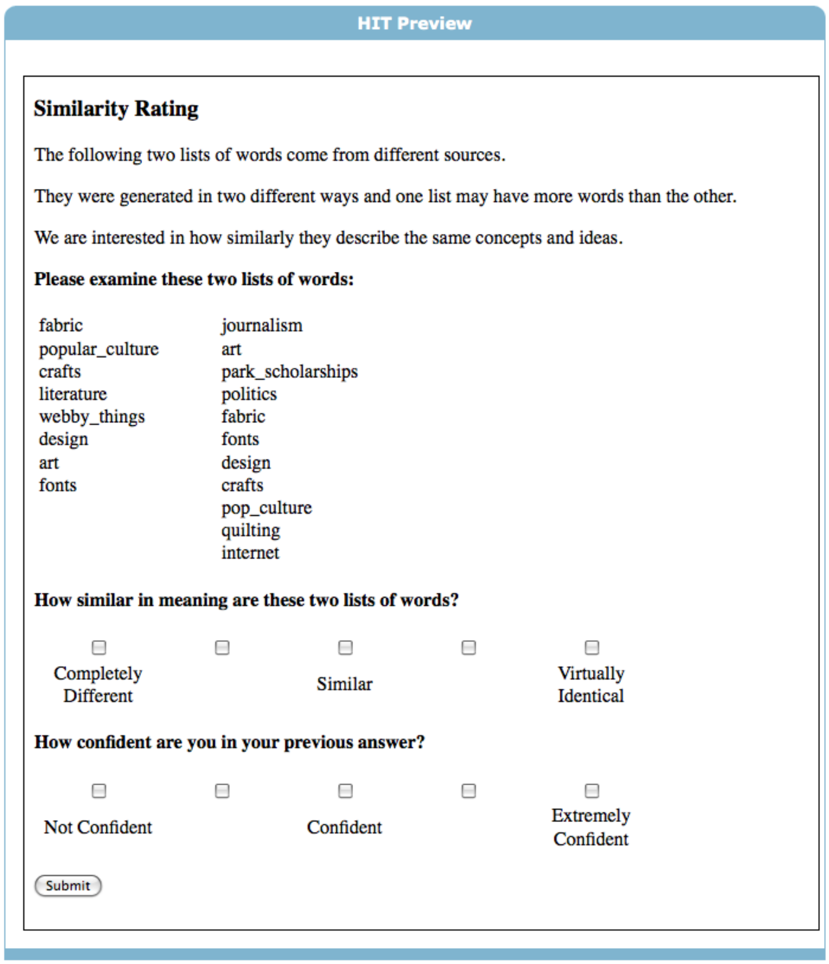

Figure 3.13: Example Similarity HIT: An example of the Human Intelligence Task (HIT) that would be presented to the Mechanical Turk worker. Completion of this task would be worth $0.02.

3.6.4.2 Algorithmically Judged

The second set of similarity scores will be computationally generated (see Figure 3.14) based on an algorithm (Equation 3.1) defined by Mihalcea et al (Mihalcea et al., 2006). The resulting similarity scores will be in the range [0..1]. The original lists of raw tags will be *sense disambiguated* and then compared against one another.

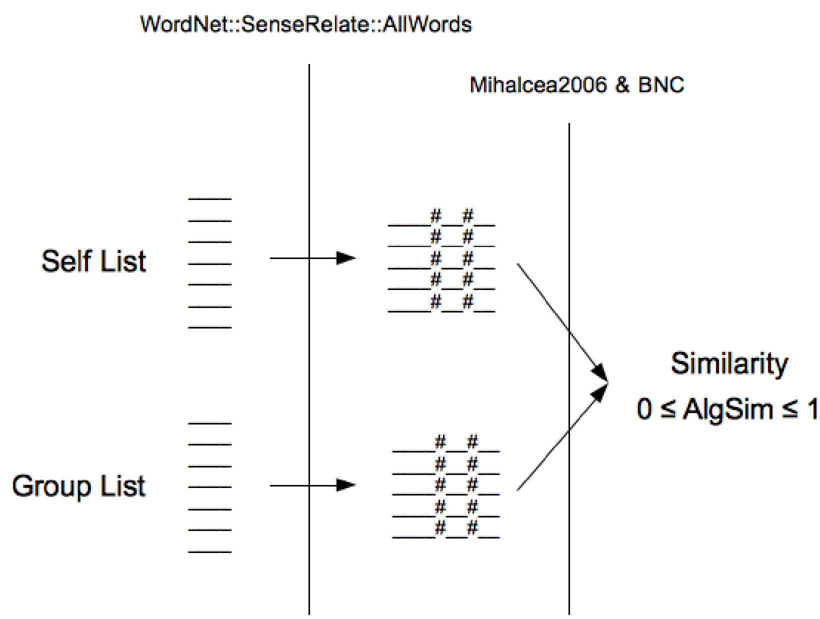

Figure 3.14: Algorithmic Similarity Rating Model – Two lists of words are 1) sense disambiguated using WordNet::SenseRelate::AllWords and then 2) compared using Mihalcea2006 (Equation 3.1) giving a Similarity score in the range [0..1].

This method does not take into consideration the word order or "sentence" structure like more recent methods (Liu et al., 2008). As sets of tags have no syntactic structure or order, Mihalcea is appropriate for this task.

The WordNet database is used to calculate similarity scores between two single words (Fellbaum, 1998) and accessed through the WordNet::Similarity[1] and WordNet::SenseRelate::AllWords[2] perl packages (Pedersen et al., 2004; Pedersen and Kolhatkar, 2009). The inverse document frequency (*idf*) of a word is calculated from the 100M word sample in the British National Corpus[3] (BNC, 2007).

Tags that are not found in the WordNet database are dropped from analysis. Tags that cannot be sense disambiguated with confidence default to the first numbered gloss, or definition, of the word. Tags that are sense disambiguated but then not found in the BNC are set to have an *idf* equal to that of the highest *idf* otherwise seen.

The Self list is processed as-is; each word has equal weight and all serve as inputs into the model. The Group list is processed through the model in two different ways. First, all the words from the group list serve as inputs, but unweighted. Second, the group list is truncated to only contain the words with a weight of two or greater. These words are then unweighted and serve as the inputs into the model.

$$AlgSim(A,B) = \frac{1}{2}\left(\frac{\sum\limits_{w \in \{A\}} (maxSim(w,B) * idf(w))}{\sum\limits_{w \in \{A\}} idf(w)} + \frac{\sum\limits_{w \in \{B\}} (maxSim(w,A) * idf(w))}{\sum\limits_{w \in \{B\}} idf(w)}\right) \tag{3.1}$$

Equation 3.1 takes each word in set A and finds the most similar word in set B (represented by $maxSim(w, B)$) and then multiplies by the information

[1] http://wn-similarity.sourceforge.net/
[2] http://senserelate.sourceforge.net/
[3] http://www.natcorp.ox.ac.uk/

content of that word (represented by $idf(w)$). This summation is normalized across the information content of the entire list ($\sum_{w \in \{A\}} idf(w)$). After each list is compared one to the other, the similarity values are averaged for the final *AlgSim* value.

3.7 Analysis

In order to address the research questions stated at the end of Section 2.6, I will conduct the following analysis.

Question	Hypothesis	Dataset(s)	Analysis
R1a - Similarity	Increasing	Lists and Similarity	Mechanical Turk and Algorithm
R1b - Convergence	Yes	Lists and Similarity	ANOVA
R2a - Comfort	Increasing	Survey and Interviews	ANOVA, Content Analysis
R2b - Confidence	Improved	Survey and Interviews	ANOVA, Content Analysis
R2c - Usefulness	–	Survey and Interviews	Content Analysis

Table 3.1: Mapping of Research Questions, Hypotheses, Data, and Analyses

Research Question 1 will be addressed with the use of the Similarity datasets coming from Mechanical Turk and the Algorithmic Similarity ratings. I will plot these values against time (Round) and expect to see the value increase. I expect the rate of change to slow over time after an initial jump in similarity ratings from round 1 to round 2. An analysis of variance (ANOVA) between each round will allow me to determine the significance of the changes over time. Additionally, this analysis can be performed at the group level and then again at the entire experiment level.

Research Question 2 will be addressed primarily with the responses to the survey and the interviews. I suspect to hear a variety of perspectives on the reasons this tool creates uncertainty and suspicion with regards to the participants level of control of what they view as their personal information. I think that participants will come to realize the contextualized nature of this medium of communication and that it provides a level of information that is not otherwise being captured somewhere else. With regards to confidence, CAT can provide a sanity-check on what an individual thinks about someone's areas of expertise. With iteration and continued use, I think confidence that the system is providing a unique service will increase.

If the participants feel that they learned about themselves or about others, then there must be some value in a system like this. Participants will come to trust that they are getting good information from a system where everyone has input. Determining whether they felt it was worth the time and energy that they devoted to interacting with the system, or whether it outweighed the potential abuses of a system like this remain to be seen. Lastly, insights into whether this could be a recommended tool for others will help drive development and further research into how a tool with third-party input can be made to feel comfortable and safe.

I plan to compare the results from the pre-test and post-test with ANOVA. I plan to compare and contrast the interview data through content analysis and coding of responses.

3.8 Potential Ramifications

If this type of methodology and analysis can be shown to be effective in the physical world, where identity is more stable and communication channels more rich and varied, perhaps it would also work in a mediated space (an online forum, gaming, or with remote workers). When identity is more malleable and easier to manipulate, a system that can provide some infrastructure and persistence could prove very useful.

When building systems that depend on expertise tagging data for input, another potentially exciting property would be the ability to "quiet" the input from those who do not meet a certain "threshold" of knowledge in an area. If a participant (human or software agent) is not deemed knowledgeable enough on a particular topic of interest, then their input could be programmatically ignored or filtered by others. Less distractions lead to much higher quality discussions among those who know what they are talking about.

Additionally, on the other hand, extra value could be given to those who *do know* what they are talking about. In an election, or key decision-making period, someone who the group deems knowledgeable in a certain domain may be routed certain questions, awarded extra votes, or have a weighted opinion counted in some other way, again, automatically or programmatically. Decisions do not have to be arrived at democratically. Most decisions in the real world are not made with equal representation.

Last, as a practical matter, organizations could use this method for deciding who to have work together when forming teams, conferences could use this method to help decide how to distribute reviewing assignments for posters and papers, and new hires into a group or company could use this system to acclimate themselves into the culture by quickly knowing who best to ask when they have questions.

Bibliography

Abecker, A., Bernardi, A., Hinkelmann, K., Kühn, O., and Sintek, M. (1997). Towards a well-founded technology for organizational memories. *American Association for Artificial Intelligence Technical Report*, SS-97-01. 11, 13

Ackerman, M. S. and Malone, T. W. (1990). Answer garden: A tool for growing organizational memory. *Conference on Supporting Group Work*, pages 31–39. 14

Balog, K., Azzopardi, L., and de Rijke, M. (2006). Formal models for expert finding in enterprise corpora. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 43–50, Seattle, Washington. 14

Balog, K., Azzopardi, L., and de Rijke, M. (2009). A language modeling framework for expert finding. *Information Processing Management*, 45(1):1–19. 13, 14

Balog, K. and de Rijke, M. (2007). Determining expert profiles (with an application to expert finding). In *Proceedings of the 20th international joint conference on Artifical intelligence*, pages 2657–2662. 14

Balog, K. and de Rijke, M. (2008). Non-local evidence for expert finding.

In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management*, pages 489–498, New York, NY, USA. ACM. 14

Becerra-Fernandez, I. (2000). The role of artificial intelligence technologies in the implementation of people-finder knowledge management systems. *Knowledge-Based Systems*, 13(5):315–320. 15

Becks, A., Reichling, T., and Wulf, V. (2004). Expertise finding: Approaches to foster social capital. In Huysman, M. and Wulf, V., editors, *Social capital and information technology*, pages 333–354. MIT Press. 13

BNC (2007). *British National Corpus*, volume 3. Oxford University Computing Services on behalf of the BNC Consortium. 50

Boje, D. M. and Murnighan, J. K. (1982). Group confidence pressures in iterative decisions. *Management Science*, 28(10):1187–1196. 24

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117. 19

Callahan, S. (2006a). Techniques for expertise location. *Anecdote*. 15

Callahan, S. (2006b). Want to manage tacit knowledge? communities of practice offer a versatile solution. Technical report, Anecdote. 12

Campbell, C. S., Maglio, P. P., Cozzi, A., and Dom, B. (2003). Expertise identification using email communications. In *Proceedings of the twelfth international conference on Information and knowledge management*, pages 528–531. 14

Choo, C. W. (1996). The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *International Journal of Information Management*, 16(5):329–340. 6, 7, 17

Craswell, N., Hawking, D., Vercoustre, A.-M., and Wilkins, P. (2001). P@noptic expert: Searching for experts not just for documents. In *Ausweb 2001*. 14

Dalkey, N. C. and Helmer, O. (1963). An experimental application of the delphi method to the use of experts. *Management Science*, 9(3):458–467. 23

Davenport, T. H. (1997). If only hp knew what hp knows. *Ernst Young Center for Business Innovation*, pages 20–25. 15

Dervin, B. (1983). An overview of sense-making research: Concepts, methods, and results to date. 4

Dieng, R., Corby, O., Giboin, A., and Ribiére, M. (1999). Methods and tools for corporate knowledge management. *International Journal of Human-Computer Studies*, 51(3):567–598. 12

Downs, A. (1957). *An Economic Theory of Democracy*. Harper and Row. 3

Duguid, P. (2005). “the art of knowing”: Social and tacit dimensions of knowledge and the limits of the community of practice. *The Information Society*, 21(2):109–118. 11

Ehrlich, K. (2003). Locating expertise: Design issues for an expertise locator system. In *Sharing Expertise: Beyond Knowledge Management*. MIT Press. 11, 13

Engelbart, D. C. (2004). Augmenting society’s collective iq. *Keynote, Hypertext 2004*. 10

Farrell, S. and Lau, T. (2006). Fringe contacts: People-tagging for the enterprise. In *WWW ’06: Proceedings of the 15th international conference on World Wide Web*, volume Collaborative Web Tagging Workshop. 15

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books. 50

Fitzpatrick, R. B. (1999). The community of science, inc. *Medical Reference Services Quarterly*, 18(3):57–63. 14, 15

Fitzpatrick, R. B. (2001). The community of science, inc., part 2. *Medical Reference Services Quarterly*, 18(4):33–38. 12, 15

Gilovich, T. (1987). Secondhand information and social judgement. *Journal of Experimental Social Psychology*, 23:59–74. 5

Golder, S. A. and Huberman, B. A. (2005). The structure of collaborative tagging systems. Technical report, Information Dynamics Laboratory, HP Labs. 20

Helmer, O. and Rescher, N. (1959). On the epistemology of the inexact sciences. *Management Science*, 6(1):25–52. 23

Hsu, C.-C. and Sandford, B. A. (2007). The delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation*, 12(10). 23

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632. 19

Kühn, O. and Abecker, A. (1997). Corporate memories for knowledge management in industrial practice: Prospects and challenges. *Journal of Universal Computer Science*, 3(8):929–954. 12

Lamont, J. (2003). Expertise location and the learning organization. *KMWorld*. 11, 12

Latour, B. and Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press. 20

Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press. 11

Ling, C. W., Sandhu, M. S., and Jain, K. K. (2009). Knowledge sharing in an american multinational company based in malaysia. *Journal of Workplace Learning*, 21(2):125–142. 11

Linstone, H. A. and Turoff, M. (1975). *The Delphi Method: Techniques and Applications*. Addison-Wesley, London. 23, 24

Liu, X.-Y., Zhou, Y.-M., and Zheng, R.-S. (2008). Measuring semantic similarity within sentences. In *Proceedings of the 7th International Conference on Machine Learning and Cybernetics*, pages 2558–2562, Kunming. 49

Lowenstein, M. (2009). Offline social word of mouth influence on brand decision-making more frequent and more powerful than online social media. *The Harris Poll, Harris Interactive*. 11

Luo, L. and Wildemuth, B. M. (2009). Delphi studies. In Wildemuth, B. M., editor, *Applications of Social Research Methods to Questions in Information and Library Science*, chapter 10. Libraries Unlimited, Westport, CT. 24

Lupia, A. (1994). Shortcuts versus encyclopedias: Information and voting behavior in california insurance reform elections. *The American Political Science Review*, 88(1):63–76. 3

Lutters, W. G., Ackerman, M. S., Boster, J., and McDonald, D. W. (2000). Mapping knowledge networks in organizations: Creating a knowledge mapping instrument. In *Americas Conference on Information Systems*. 14

Martin, B. (2008). Knowledge management. *Annual Review of Information Science and Technology*, 42. 7

McDonald, D. W. (2001). Evaluating expertise recommendations. In *Proceedings of the ACM 2001 International Conference on Supporting Group Work (GROUP)*, pages 214–223. 14

McDonald, D. W. and Ackerman, M. S. (1998). Just talk to me: A field study of expertise location. In *Computer Supported Cooperative Work*, pages 315–324. 14

Mihalcea, R., Corley, C., and Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, pages 775–780. AAAI. 49

Nelson, L. H. (1993). Epistemological communities. In Alcoff, L. and Potter, E., editors, *Feminist Epistemologies*, New York. Routledge. 20

Nonaka, I. (1991). The knowledge-creating company. *Harvard Business Review*, 69(6):p96 − 104. 3, 17

Pedersen, T. and Kolhatkar, V. (2009). Wordnet::senserelate::allwords - a broad coverage word sense tagger that maximizes semantic relatedness. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Demonstration Session*, pages 17–20, Boulder, CO. 50

Pedersen, T., Patwardhan, S., and Michelizzi, J. (2004). Wordnet::similarity - measuring the relatedness of concepts. In *Proceedings of the 19th National Conference on Artificial Intelligence*, pages 1024–1025, San Jose, CA. 50

Reeves, B. and Nass, C. (1996). *The Media Equation*. Cambridge University Press. 11

Rohrbaugh, J. (1979). Improving the quality of group judgment: Social judgment analysis and the delphi technique. *Organizational Behavior and Human Performance*, 24(1):73–92. 24

Rowe, G. and Wright, G. (1999). The delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, 15(4):353–375. 23, 24

Rowe, G., Wright, G., and McColl, A. (2005). Judgment change during delphi-like procedures: The role of majority influence, expertise, and confidence. *Technological Forecasting and Social Change*, 72(4):377–399. 21, 23

Russell, T. (2006). Cloudalicious: Watching tag clouds over time. *Joint Conference on Digital Libraries (JCDL)*, page 364. 20

Savolainen, R. (2007). Media credibility and cognitive authority. the case of seeking orienting information. *Information Research*, 12(3). 5

Scheibe, M., Skutsch, M., and Schofer, J. (1975). Experiments in delphi methodology. In Linstone, H. A. and Turoff, M., editors, *The Delphi Method: Techniques and Applications*, chapter IV.C., pages 257–281. Addison-Wesley. 24

Simon, H. A. (1957). *Models of Man: Social and Rational*. Wiley, New York. 5

Stein, E. W. (1995). Organization memory: Review of concepts and recommendations for management. *International Journal of Information Management*, 15(1):17–32. 12

Sweller, J., van Merrienboer, J., and Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3):251–296. 4

Van De Van, A. H. and Delbecq, A. L. (1974). The effectiveness of nominal, delphi, and interacting group decision making processes. *The Academy of Management Journal*, 17(4):605–621. 24

Walther, J. B., Heide, B. V. D., Hamel, L. M., and Shulman, H. C. (2009). Self-generated versus other-generated statements and impressions in computer-mediated communication: A test of warranting theory using facebook. *Communication Research*, 36(2):229–253. 21

Walther, J. B. and Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In Knapp, M. L. and Daly, J. A., editors, *Handbook of interpersonal communication*, pages 529–563. 21

Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In Mullen, B. and Goethals, G. R., editors, *Theories of group behavior*, pages 185–208, New York. Springer-Verlag. 17

Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority*. Greenwood Press, Westport, CT. 4, 5, 19

Yamim, D. (1996). Expert finding systems for organizations: Domain analysis and the demoir approach. In *ECSCW 1999 workshop: Beyond knowledge management: Managing expertise*, pages 276–283. 14

Zhang, J., Tang, J., and Li, J. (2007). Expert finding in a social network. In *Advances in Databases: Concepts, Systems and Applications, DASFAA 2007*, Lecture Notes in Computer Science, pages 1066–1069. 14