Impression Formation in Corporate People Tagging

Daphne R. Raban and Avinoam Danan
Graduate School of Management
University of Haifa, Israel
draban@univ.haifa.ac.il; noam.da1@gmail.com

Inbal Ronen and Ido Guy
IBM Research Lab
Mt. Carmel, Haifa 31905, Israel
{inbal, ido}@il.ibm.com

ABSTRACT
This research explores the relationship between self-presentation and perception by others as manifested explicitly through the use of tags in a people tagging system. The study provides insights relevant for the organizational context since it is based on a system implemented within IBM. We developed a detailed codebook and used it to categorize 9,506 tags assigned to a sample of taggers. Our analysis examines the use of self tags versus social tags (assigned by others) across different categories and sub-categories. While overlap exists, self tags tend to be more factual describing technology expertise, social tags augment the individual tags by adding a personal dimension.

Author Keywords
Impression management, presentation of self, perception of others, people tagging.

ACM Classification Keywords
H.5.3 Group and Organization Interfaces: Computer-supported cooperative work.

General Terms
Design, Experimentation, Human Factors

INTRODUCTION
Internet technology offers users new tools and multiple opportunities for expressing themselves online. Social information technologies allow users to interact, collaborate, and socialize with each other in new and diverse forms such as blogs, social networks, wikis, content sharing applications and more.

As the web is transforming from a content focus to an increasingly social arena, the importance of self-presentation and the perception by others becomes evident. Even President Obama who is famous for efficiently using the social web in his 2008 election campaign alluded to the importance of self-presentation by his remark to high school students: "*Be careful what you post on Facebook*."

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Moreover, impressions are perceived fast and sometimes remain regardless of newer information that is received about people. This has been studied in the marketing literature as the halo effect, in decision making as the confirmation bias and in the usability and visual appeal of web sites [12]

Impressions online can be based on implicit or explicit cues through which people present themselves on the web. An explicit form of describing people is a people tagging system. In this system people describe themselves and their peers by the use of tags. This presents a unique opportunity to study explicit self-presentation and perception by others using real-life data.

Online impressions have been studied to some extent as we describe later, however, we wish to shed new light by examining the following dimensions:

1. Studying self-presentation and concurrently the perception by others.
2. Focusing on the organizational environment.
3. Analyzing a system that offers an explicit form of self-presentation and perception.

This paper will briefly describe the major domains that are related to the research: people tagging, self presentation and perception of others. We then outline a research question, and the research method selected for the current work. Subsequently, we provide results based on a sample of 9,506 tags by a sample of 281 most active taggers, analyze their implications and suggest future research directions.

PEOPLE TAGGING
In online information systems terminology, a tag is a non-hierarchical keyword or term assigned to a piece of information in order to indicate identity, mark ownership, contextualize, etc. Tags enrich the information about an item and can be used for later retrieval. Since tagging is usually done by end-users, the result is the creation of a web of human topical interest.

Tags can be chosen from a controlled vocabulary or they can be free text assigned by the user. When free text is used, the resulting metadata can include homonyms, synonyms, spelling mistakes and errors which can lead to inappropriate connections between items and insufficient outcomes for information searches. On the other hand, with free text the user can use terms he thinks are appropriate to

describe or help him recall information without the burden of selecting a category from a known taxonomy. Free text also allows for a dynamic update of the vocabulary, maintaining its relevance.

It is worth noting that tags are often presented in the form of tag clouds which are a graphical presentation of a weighted list of words. Is word frequency the only valuable way to organize tags? This practical design question may benefit from the outcome of the present study.

The special case where the resources to tag are people, is referred to as people tagging. The uniqueness of people tagging in contrast to other tagging applications can be derived from the Social Relations Model (SRM) [11], which argues that perception of people is different from perception of objects in four ways: it is reciprocal, it invokes thoughts about how the other will receive what one has told him (meta-perception), other perception is closely linked to self perception, and finally, perception may change over time. From the individual's perspective, people tagging systems can be described as two-sided, enabling use for acting and reacting, posting or replying [16], therefore, it is an opportunity for a field study of the SRM.

In their study, Ehrlich and Cash claimed that people in an enterprise can provide valuable expertise to solve problems [4]. Knowledge work (e.g expertise, unique knowledge) often involves finding opportunities to contribute to collaborative work, and one way, possibly the simplest way, of finding those opportunities is to "advertise" one's skills to other members of the organization through the use of tags [14].

In fact, the most common use of people tagging systems is for contact management. People use the application to arrange their friends and colleagues in a social network system [6]. Beyond contact management, people tagging enables the formation of an organizational community which collectively maintains each others' interests and expertise. This expanded profile is searchable and available for later recall about employees associated with a project or a needed skill or knowledge.

SELF PRESENTATION

In order to be socially-accepted or cultivate certain reputation or recognition, people tend to try and control the information they present about themselves in one-to-one encounters as well as in group meetings or before large audiences. Goffman distinguished between two kinds of expression of the self: expressions one gives, and expressions one gives off [8]. The first involves relatively easily controlled, presumably intentional expressions, conveyed through traditional verbal communication [15, 23]. The other kind, implicit expression, is considered to be more theatrical and contextual, nonverbal, and presumably unintentional [15].

Any situation of appearance before others involves self presentation and induces people to express themselves explicitly and implicitly. By attempting to control or manage the impression others receive people will calculate behavior so that the audience will believe what they see [8].

Self Presentation Online

Self presentation, also referred to as impression management, implies that every individual performs a certain role while appearing in public [23]. The internet and organizational systems provide several ways for online self presentation such as personal home pages, the profile one accumulates for oneself on a variety of online systems, blogs, email, instant messaging, social networks, dating sites and intranet tools for contact management within organizations.

In each of these tools, the participant chooses how to introduce himself, how to manage his impression, and what is his preferred level of participation. A user can introduce himself by choosing nicknames and perhaps forfeiting age, gender, workplace, or other cues [18]. The web offers opportunities to expose multiple senses of self or personas in different contexts [24] to the point of playfulness or fantasy as forms of misrepresentations. There is a tension between idealizing the self and presenting truthfully, however, individuals usually know how to balance their desire for self-promotion with their need for accurate self-presentation [5][1][13].

Anticipation of a future meeting has been shown to favor truthful self presentation [7]. While online self-presentation is more malleable and subject to self-censorship than face-to-face self-presentation [26], under certain conditions the online medium may enable participants to express themselves more openly and honestly than in face-to-face contexts [9]. An investigation of truthfulness of presentation was done in a dating web site where people used single adjectives, somewhat similar to the notion of tagging, to describe themselves and their desired partners. The results showed that interacting over the Internet led to a more truthful self presentation leading to a better match [27]. Thus the internet serves as a social lubricant by delivering a clear self-presentation which meets the expectations of others.

In a work environment, the tension between the need for self-promotion and truthful representation is expected to be minimal because people participate in the system as part of professional community building. If they describe themselves as having a certain skill they do so out of confidence in their ability in this area. We see no particular motivation to idealize one's skills in an organization.

Overall, we believe that in organizations, self-presentation will tend to be less personal, more professional – less about appearance or age and more about knowledge and skills, and similar to the dating studies cited here, it will be truthful. The purpose of organizational self-presentation is

social in a work-related sense – to find like-minded people or locate experts to help in promoting a project [2].

People tagging is uniquely reciprocal, as users can tag themselves, others or both. While people share information as a form of self presentation, other individuals may perceive them differently than what they present and tag them using other terms.

PERCEPTION OF OTHERS

Perception of others online is distinct from face-to-face situations. Rouse and Haas [19] pointed out the main factors that explain the differences between online and face-to-face interactions: 1. Physical appearance has a less meaningful effect. 2. People may behave differently online especially when anonymized. 3. In online communication there may be a heightened level of ambiguity due to the lack of vocal inflection and facial expression.

Online perception can be either synchronous or asynchronous. A person creates an impression by his digital artifacts such as writing, photos, selection of materials for inclusion in a web page or personal profile [21]. Impressions can form through synchronous conversation but also by a visit to a homepage or profile page. As the interaction progresses more information accumulates, verbal and non-verbal cues are taken into consideration and affect the impression [20]. Perception of others online is constructed according to the explicit and implicit cues observed. In order to make a positive impression a person will carefully select the explicit cues to manifest. The online perceiver is aware of that and is likely to notice the implicit cues as well.

An example of such a communication was identified in Q&A websites where both the asker and the answerer tried to make a positive impression on each other; the asker for getting valuable answers and the answerer for being rewarded positive feedback and a monetary tip for his answer [17]. The present study focuses on explicit cues.

Prior research suggests that explicit cues will deliver a reliable representation of a person especially when a future meeting is anticipated [5, 9]. Why, then, is tagging practiced both by the person himself and by his perceiver? Why is there a need to tag a person beyond his own self-description? According to the SOKA model (Self Other Knowledge Asymmetry) differences between self and other-perception derive from informational differences in perspective (i.e., the salience of overt vs. covert aspects of a person) and motivational significance (such as the need for self-enhancement) [25]. Thus, self descriptions are more accurate than others for internal, unobservable, less evaluative traits (e.g. neuroticism), and others are slightly more accurate than the self for external, observable and evaluative traits (e.g. intelligence). Therefore, the perspectives offered by the actors themselves vs. those offered by their peers are expected to differ and complement each other. Corporate people tagging offers a window for examining the degree of similarity or difference in the explicit presentation and perception of employees using the tagging system.

RESEARCH QUESTION

People tagging offers a unique opportunity to study explicit self-presentation and perception by others by analyzing tags. Tags are explicit cues because users describe themselves or others using individual words or short phrases, without implicit effects such as writing style, manners, or use of personal pronouns. We use the tags in order to examine both sides of the coin: self-presentation and perception by others, and to learn whether and how they complement each other.

Using the database of a people tagging application implemented in a large enterprise we explore the following question: What terms do people use to tag themselves (presentation of self) compared to terms used by others (perception by others)? To what degree are the self descriptions similar to or different from the tags assigned by others? Do self tags converge with incoming tags or is a larger number of tags associated with a more diverse description?

METHOD

The data was harvested from an organizational profile application within IBM which included a people tagging system based on free-text tags. It represents a period of three years during which the application was deployed on the organization's intranet. Each employee in the application had a profile page which exposed the tags used by the employee to tag himself (referred to as ***self tags***) and a tag cloud of the tags assigned to him by other employees (referred to as ***social tags***). Additionally, profile information was shown such as the person's name, photo, job title, department and others. The same social tag could be applied several times by different persons. A self tag could be applied only once. Employees used the people tagging application both to tag themselves (self presentation) and to assign tags to others (perception of others). Figure 1 shows an anonymized screenshot of a typical profile in the application. For the present research the preferred methodology is content analysis which allows an understanding of the meaning of the tags.

The people tagging application kept each tag with the id of the tagged person, the tagger and the date of the tagging event in a database. A tag was stored normalized, such that the terms of a multi-term tag were separated by a hyphen. We extracted the tags related to a person by querying this database through SQL.

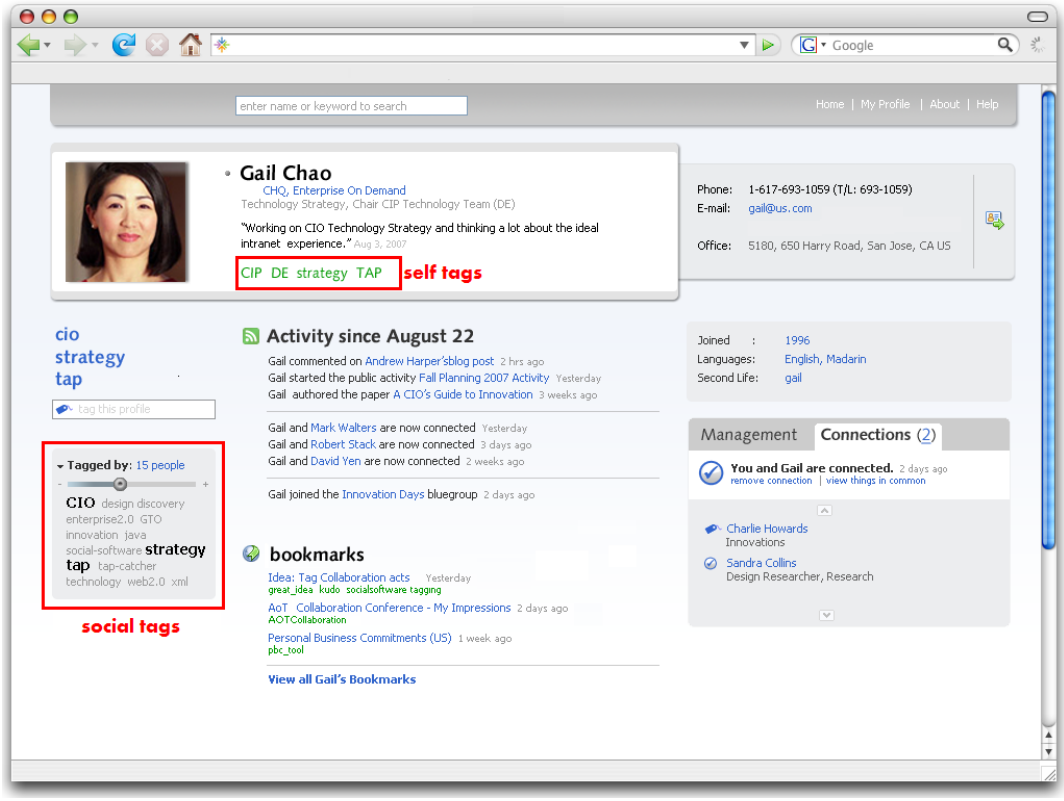

Figure 1. Profile Page in People Tagging application (anonymized)

Variables

Self Tags: tags assigned to a user by himself, representing the presentation of self.

Social Tags: tags assigned to a user by other users, representing the perception by others. Each unique social tag was counted once even if it was used by more than one person to tag the same peer.

Sample

The complete data included the tagging activity by 62,332 system participants. As in many web-based applications the distribution of tags followed a power law function; hence, random sampling would not be representative of the full data. The focus of this study is on the comparison of self tags to social tags; therefore, both types of tags must have significant counts. Based on the frequency of self and social tags we decided that a minimum of 10 in each tag set would be the cutoff for the sample. This resulted in a group of 281 users who had a total of 9,506 tags.

Content Analysis

Kaid and Wadsworth [10] suggest seven steps for implementing content analysis:

1. Formulate the hypotheses or research question.
2. Select the sample to be analyzed.
3. Define the categories to be applied.
4. Outline the coding process and train coders.
5. Implement the coding process.
6. Determine reliability and validity.
7. Analyze the results from the coding process.

Codebook development: The heart of content analysis is the coding process. We prepared a codebook by assigning the tags of 25 random system users (not from the 281 people) to three coders. Each coder described the tags according to his understanding. Then a meeting was held to form an agreement on the coding categories. Having organized the tag descriptions we identified 3 main categories: technology, environment, and individual. Every main category contains 6 sub-categories that represent the fine-tuned aspects of the category (see Table 1).

Inter-coder reliability: This is a measure of whether independent judges make the same coding decisions while characterizing the data. Without reliability of the coding, the analysis cannot be trusted [22]. We conducted inter-coder reliability for the three coders by using a new random sample of 30 participants, not from the 281 most active users. For each participant the self tags and the social tags were coded. For the first 9 people all three coders worked jointly to reach an agreed understanding of the use of the codebook. Then each coder coded independently the self and other tags for the remaining 21 people. A total of 175 self tags and 434 social tags were coded to test inter-coder reliability. Krippendorf's alpha for the three coders was 0.75 for the self tags and 0.73 for the social tags for the sample of 30 participants. Both values are above the threshold of 0.7 needed in order to determine the coding process as reliable.

Main Cat.	Sub-category	Description	Examples
Technology	Information Technologies	Information technologies and systems	linux, android, ooa, eclipse-plugins
	Internet Technologies	Internet programming, technologies and systems	web20, secondlife, atom, javascript, 3dinternet
	Theme	broad topic	collaboration, knowledge-management, legacy-transformation
	IBM Product	products developed and sold by IBM	websphere-commerce lotus-notes lotus-domino websphere-portal sametime, quickr
	External Product	Products not developed or sold by IBM	photoshop, mac, iphone
	Project	Projects (not products) of IBM	project-zero, p-vista oasis, bluepedia
Environment	Group Affiliation	Groups such as conferences and internal or external events	cscw2011, web20forbiz, helsinki-2007, class-of-78, Vienna-team
	Organizatio-nal Group	Departments and organizational groups at IBM	research, sales, services, corporate
	Country	Country names	Norway, Canada
	City	City names	Hamburg, Cambridge
	Customer	Customer and partners of IBM	Philips, Miele, Rails, Citybank
	Organization	Organization names which are not customers or partners of IBM	BMW, Cisco, Apple, Daimler, Teva
Individual	Name	Name, Last Name, Nickname	Bob, Laszlo
	Training	Academic or professional degree, training, institution	MBA, Certified-IT-Specialist, Miami-University
	Job Description	Description of job responsibility	architect, presales, manager, designer
	Special Skill	work related skills	hacker, mentor, demo-guru, speaker, bilingual
	Hobby	Non work related hobby or interest	Skiing, karaoke, photography, rocker, piano, soccer-fan
	Personal Adjective	personality related characteristics	leadership, traveler, creative, funny, supergenious, storyteller, recycler

Table 1. Codebook categories, sub-categories, and examples.

For the sample of 281 highly active participants, we applied the coding retrieved in the previous step and extracted only those tags that had not been coded before. These were distributed among the coders for independent coding. All the coded tags were then aggregated to assign the coded tags to each system participant in the sample.

RESULTS

Tag Volumes and Ratio

We first inspect the volumes of self and social tags across the three main categories. On average, each of the 281 taggers was tagged with 16.26 self tags and 17.57 social tags, leading to an overall of 4569 self tags and 4937 social tags inspected in our study.

	Mean Self Tags	Mean Social Tags	Ratio of Means (self/social)
Technology	9.15	8.02	1.14
Environment	4.0	5.03	0.8
Individual	3.11	4.52	0.69

Table 2. Mean tags per user in the 3 main categories (N=281).

Table 2 shows the mean number of self and social tags. The rightmost column in Table 2 highlights the overall ratio between self and social tags. The most active category is Technology, which is also the only category where the volume of self tags is higher than social tags. While taggers sharply prefer to tag themselves with Technology tags (over 56% of the self tags belong to Technology), this tendency is milder for social tags (less than 46% of the social tags belong to Technology). On the other hand, while only 19.2% of the self tags are Individual, 25.7% of the social tags are Individual.

A correlation analysis shows a statistically significant correlation between tagging in one main category and tagging in other main categories. This relationship is stronger for social tagging ($r=.45$ to $.50$, $p<.01$) than for self tagging ($r=.19$ to $.25$, $p<.01$).

Table 3 lists the means and ratios across the 18 sub-categories with noteworthy observations in boldface. For Technology sub-categories, the most frequent sub-categories are Theme and IBM Product. The ratio of the External Product sub-category is exceptionally high, while the ratio for Project is especially low. It seems that taggers prefer to associate others with internal projects, probably for personal bookmarking purposes [6], than to associate themselves with such projects. Interestingly the ratio difference between IBM Product and External Product suggests that the latter has stronger preference for usage as a self tag.

	Mean Self Tags	Mean Social Tags	Ratio of Means (self/social)
Information Tech	0.85	0.78	1.09
Internet Tech	1.93	1.67	1.16
Theme	**2.54**	**2.16**	1.18
IBM Product	**3.03**	**2.64**	1.15
External Product	0.37	0.17	**2.15**
Project	0.43	0.6	**0.72**
Group Affiliation	**1.57**	**1.93**	0.81
Org Group	**0.89**	**1.5**	**0.59**
Country	0.32	0.36	0.89
City	0.35	0.41	0.84
Customer	0.43	0.32	**1.34**
Organization	0.44	0.5	0.88
Name	0.05	0.13	0.41
Training	0.18	0.13	1.38
Job Description	**1.09**	**1.64**	0.66
Special Skill	**0.74**	**1.06**	0.7
Hobby	0.32	0.32	**1.0**
Personal Adj	**0.72**	**1.23**	0.59

Table 3. Mean tags per user across the 18 sub-categories.

Among the Environment sub-categories, Group Affiliation is the most frequently used, followed by Organizational Group. The latter has the lowest self-to- social ratio within the Environment category. This implies that people are less passionate to tag themselves with their organizational affiliation and aim to tag others with such affiliation more frequently. The difference in the ratio between these two sub-categories may result from the person himself being more knowledgeable than his peers about his own external group affiliations. Customer is the sub-category with highest ratio. We again observe that tags of an external entity or affiliation are more frequently used as self tags than as social tags.

As for the Individual category, Job Description is the most frequently used sub-category, followed by Personal Adjective and Special Skill. Name is used very infrequently, mostly as a social tag, hinting that taggers view it as redundant information. Training is also quite infrequent, used more often as a self tag. The low activity around academic or professional training may imply a greater focus on actual work experience.

The relatively low ratio of Personal Adjective may reflect that some employees feel uncomfortable tagging themselves with such subjective tags. Interestingly, Hobby has exactly the same number of self and social tags. The relatively high ratio implies that people feel more comfortable to testify on themselves with respect to hobbies than with respect to other subjective traits, such as adjectives or special skills.

Overall, the three most frequent sub-categories for self tags are related to Technology. They represent over 46% of the total self tags. The three leading sub-categories for social tags represent only 38% of the total social tags, indicating more diversity across social tags.

Overlap Analysis

We next set to explore the agreement between self and social tags per user across the different categories and sub-categories. We define *overlapping tags* as tags that

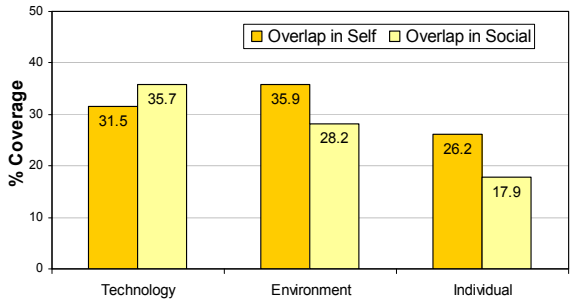

Figure 2. Overlap between self and social tags across the 3 main categories.

appear both as self and social tags for the same user. We examine two measures: (1) *overlap in self tags*, defined as the percentage of overlapping tags out of all self tags, averaged across all 281 taggers, and (2) *overlap in social tags*, defined as the percentage of overlapping tags out of all social tags, averaged across all taggers.

Figure 2 shows these two measures for the three main categories. The left bars represent the overlap in self tags and the right bars represent the overlap in social tags. Inspecting both Table 2 and Figure 2 shows that a lower mean number of tags in one of the main categories is associated with a lower overlap in the same category. The Technology category has the highest number of tags and the highest rate of overlap, while the Individual category has the lowest number of tags and the lowest overlap.

Figure 3 shows the overlap analysis for the sub-categories across the 3 main categories. For the Technology category, the two sub-categories with highest overlap are IBM Product and Internet Technology. The sub-category with least overlap is Non-IBM Product.

For the Environment category, most noticeable is the high overlap in self tags for Organizational Group. The lowest overlap was associated with Group Affiliation.

Overlap in the Individual category is generally lower than in the other main categories. Note that the overlap for Name and Training is based on especially low volume of tags (see Table 2) and should thus be treated with a grain of salt. Special Skill has the highest overlap between self and social tags. The overlap in self tags for Job Description is lower than for Personal Adjective and for Special Skill. This is an interesting finding: people's presentation of self regarding their unique skills agrees with the perception of others more often than for their job description. While hobbies have identical volumes of self and social tags, the overlap of their content is only 18%. It seems that the hobbies people want to be associated with are not the ones others associate them with.

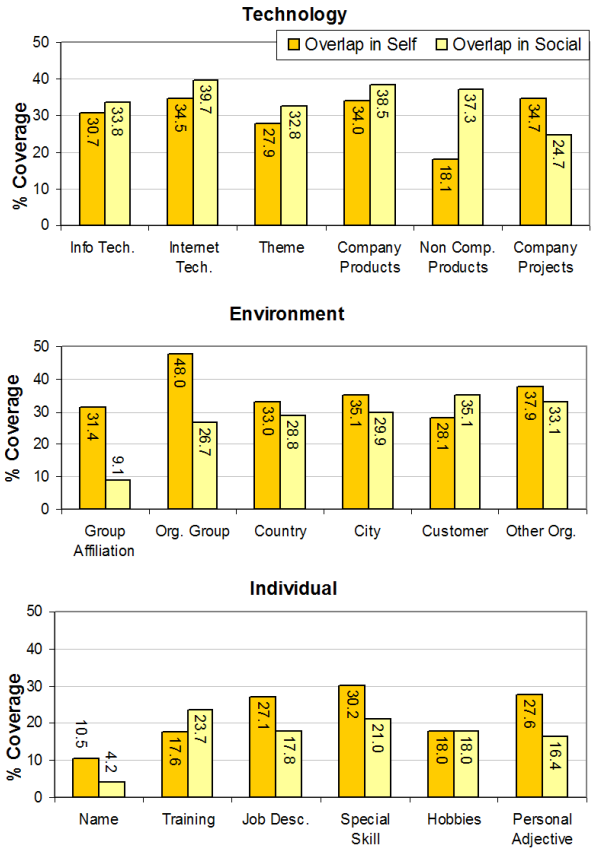

Figure 3. Self and Social Overlap across the 18 sub-categories.

Tag Popularity Analysis

In this part of the analysis we focus on the most popular tags in each category and sub-category. Table 4 shows the three most popular tags in each of the three main categories. The right section of the table indicates the overlap between the k most popular self and social tags, for k=1,5,10. We denote by *ovl@k* the portion of tags that are amongst the k most popular self tags as well as the k most popular social tags. For example, an ovl@5 of 0.6 indicates 3 overlapping tags within the 5 most popular. Note that the binary ovl@1 directly indicates whether the most popular tag is identical for self and social tags.

Table 5 shows the most popular self and social tags, as well as their overlap, for the sub-categories. We excluded the Name and Training sub-categories since they did not have enough tags to generate rich enough popularity lists.

In general, the overlap rates for the sub-categories are high: 12 of the 16 sub-categories have an identical top tag for self and social; the rates for ovl@5 and ovl@10 range mostly from 0.4 to 0.8. Sub-categories with higher volumes of tags have higher overlap, suggesting convergence. The overlap rates also point at sub- categories with more narrow and/or well-defined vocabularies. For example, Organizational Group has very high overlap and so does IBM Product (in spite of disagreement on the top tag).

	Popular Self Tags	Popular Social Tags	ovl @1	ovl @5	ovl @10
Technology	web20 secondlife collaboration	web20 collaboration secondlife	1	0.8	0.7
Environment	lotus web204biz vuccore	lotus web204biz group2	1	0.4	0.7
Individual	blogger architect competitive	blogger architect innovator	1	0.4	0.5

Table 4. Popular tags and their overlap across the 3 main categories.

There is less agreement within the most popular Individual tags, implying that people use somewhat different vocabularies when describing self versus others with tags that tend to be more subjective. For example, consider the Personal Adjective sub-category: 'competitive' is among the top three for self tags, but is not among the top 10 for social tags, as it might be sensed as a negative perception of another person. When testified by self, it can be perceived as a more positive quality in a workplace. On the other hand, 'innovator' is the top social tag, representing a tag that one would mostly want to get from others rather than use for self tagging. There are also common adjectives for self and social, such as 'early adopter' or 'evangelist'.

DISCUSSION

This research set out to explore how people present themselves, compared to how they are perceived by their peers in a corporate setting when they are given a tool for explicit description, the people tagging system. Our analysis goes from broad to focused view. We started by looking at overall and mean tag counts in the three main categories. We, then, zoomed into the sub categories to gain a finer description of the person by himself and by his peers. Next, we employed overlap analysis to assess the degree of similarity between self and social tagging. Lastly, the focus was on the most frequently used overlapping tags.

The Broad View

The sample analyzed is the group of the 281 most active taggers with ten or more self and social tags each. By focusing on this group we ensured a sufficient description for each person in the sample. While this may not be representative of the average user, learning how the most active taggers act can be a point of reference for system designers.

Looking across the rows in the broad view presented in Table 2, the number of self tags is quite similar to the number of social tags per category. Previous work [16] showed that self tagging corresponds with more incoming, social, tags. Table 2 suggests that not only does tagging invoke tagging but also of the same category. A person who tags himself with Technology or Environment tags, is likely to be similarly tagged by others.

	Popular Self Tags	Popular Social Tags	ovl @1	ovl @5	ovl @10
Info Tech	linux java css	linux java visualization	1	0.6	0.6
Internet Tech	web20 secondlife social-networking	web20 secondlife social-networking	1	0.8	0.7
Theme	collaboration innovation social-computing	collaboration innovation work	1	0.8	0.8
IBM Product	webssphere-portal lotus-connections lotus-notes	lotus-connections sametime websphere-portal	0	0.8	0.9
External Product	twitter eclipse mac	twitter mac moleskine	1	0.6	0.4
Project	dogear bluepedia beehive	beehive bluepoints ets	0	0.6	0.5
Group Affiliation	web20forbiz vuccore blueiq	web20forbiz blueiq vuccore	1	0.8	0.8
Org Group	lotus websphere swg	lotus gbs sales	1	0.8	0.9
Country	Germany Australia Austria	Germany UK Austria	1	0.8	0.7
City	Hamburg Hursley Stuttgart	Hamburg Hursley Berlin	1	0.6	0.4
Customer	ebic goverment itil	ebic deutsche-bank siemens	1	0.2	0.2
Org	Cisco SAP Apple	Cisco Apple Xing	1	0.4	0.7
Job Desc	architect tech-sales manager	architect manager tech-sales	1	0.6	0.5
Special Skill	blogger mentor community-builder	blogger mentor 5live-speaker	1	0.5	0.6
Hobby	photography writer soccer	biker photography writer	0	0.4	0.3
Personal Adj	evangelist early-adopter competitive	innovator evangelist early-adopter	0	0.4	0.7

Table 5. Most popular tags and their overlap across 16 of the sub-categories.

Possibly at the Individual category social tagging is more intensive because people may shy away from assigning themselves personal descriptors. The self-to-social ratio supports this direction showing that people are more comfortable tagging themselves with matter-of-fact technology tags. Peers make a greater effort to augment the individual tags, adding a personal dimension to the tag cloud.

Correlation analysis confirmed that tagging in one main category was associated with tagging in other main categories. Moreover, this relationship was stronger for social tagging which suggests diversification. More tags lead to a more diverse description of a person over the three main categories, not "more of the same" tags in a specific category.

The Deeper View

The sub-categories in Table 3 help to understand what specific information is more useful to system users. Here, we see that in the Technology category the most used sub-categories were IBM Product, Theme and Internet Technology, in that order. These sub-categories are an indication of the employee's specific context (Product) as well as his general knowledge or association (Theme, Internet Technology). The self-to-social ratio in all three leading sub-categories is higher than 1, suggesting the importance of these sub-categories for the presentation of self which tends to be factual. On the other hand, perception by others is based more on the sub-categories belonging to the Environment and Individual categories. Specifically, perception is based on the sub-categories Group Affiliation, Organizational Group, Job Description, Special Skill, and Personal Adjective. For all these sub-categories the self-to-social ratio is smaller than 1 – taggers feel they need to augment these socially-oriented sub-categories more than other factual technology descriptions. This is in line with the SOKA model [25] in two ways. First, tagging is complementary by contributing different perspectives, and second, social tags tend to be more about external circumstances or character traits that a person might not think about highlighting for oneself.

Sub-categories used in the Individual category underscore what personal information is useful. It is a mixture of fairly objective information such as Job Description (e.g. programmer, architect, manager) with subjective information: Special Skill (hacker, speaker, mentor) and Personal Adjective (innovator, creative, funny). These unique descriptions add a human dimension to the system and enhance the formal, general, descriptions, with qualities unique to the person. To determine the degree of uniqueness of the descriptions we turn to overlap analysis.

Overlap in Presentation and Perception

The degree of overlap reflects the agreement between self tags and social tags per user, by considering the portion of tags that appeared as both. Overlap is an indication of common knowledge which can be viewed as part of corporate cultural capital. Overlap also indicates where we might obtain new information – categories with the lowest overlap. We should encourage users to populate those categories more.

The Technology category has the highest number of tags and the highest rate of overlap. The more people get tagged, the more overlap develops. This suggests that convergence is occurring: more tags result in more similarity in tagging in line with an earlier study [16]. This is also in line with the SRM model which predicts similarity in presentation and perception [11].

Within the Technology category, we observe high agreement on IBM Product, as opposed to low agreement with regards to Non-IBM Product. In the Environment category, there is a high agreement on Organizational Group and a low agreement on Group Affiliation. Both findings indicate that presentation of self and perception by others more often agree when it comes to organizational association, be it a product or a group, than an external one. Within the Individual category, there is surprisingly more agreement for Special Skill than for the supposedly-more-objective Job Description.

Popular Tags

The results in Table 4 indicate that for all three main categories the most popular tag is identical for self and social tags. In general, the overlap within the popular tags is considerably higher than the overall overlap, as depicted in Figure 2, indicating that both types of tags converge to a similar vocabulary. The Technology category has the highest overlap, presumably due to its higher volumes of tags, which allows for more convergence.

The overlap in the most highly used, popular, tags is very high, indicating that almost a consensus is reached in the overall usage of terms. This can be seen as kind of a 'wisdom of crowds', surfacing the organizational vocabulary or cultural capital. The agreement is reached for quite general terms such as listed in Table 5 (Linux, Web 2.0 etc.). What individuates a user is found further down the path of tags in the less popular areas of the tag list.

Another insight from this very high overlap in the top tags is that tagging is to a large extent reciprocal. For example, if a user tags himself with 'Linux' statistical probability suggests that he will be tagged by a peer with the same term and vice versa.

Study Limitations

The current study is based on a sample of the most active taggers of the people-tagging system; therefore, the results are limited to understanding the practices of this sub-population. On the other hand if we treat this group as a benchmark for the system's potential then we can identify goals for further system development. A similar study based on a representative sample of the entire population of users could help mark the starting point and suggest paths from the lower levels of activity to higher ones.

From a technical standpoint, we did not identify multi-term tags which had no explicit separation between the terms. For example, social-networking and socialnetworking were treated as different tags. No stemming or synonym identification has been done. We estimate that these issues introduce an error rate of 3% at the most so correcting them would not make a significant difference in the results presented here.

CONCLUSIONS

We studied presentation of self and perception by others as reflected through enterprise people tagging. People use tagging in critically different ways, but as a collective enterprise use it well. The most common form of self presentation is through Technology tags, mainly IBM Product, Theme, and Internet Technology. While there is considerable overlap, especially for popular tags, there is also a substantial degree of complementary tagging, bringing new information to the system.

Self presentation often relies on Technology tags and is less personal than one would expect from people describing themselves. People use less individuating tags when managing their own impression. This is filled-in to some extent by the social tags. Possibly, the work context induces this type of presentation vs. perception so it is even more important in this context to cultivate more uniqueness in impression management.

Environmental and Individual tags are generally used more for augmenting the perception of people than for self presentation. Exceptions to this rule include the Customer and Training sub-categories, which are used more for self presentation. Customer probably reflects the greater knowledge one has and wished to expose regarding the customers relevant to his work. This may not be well-known by one's colleagues. Training is typically known by the person and is used for self-advertising and not so much for identification by others.

The broad view we presented suggests diversification of tags while the overlap and popular tags analyses point toward convergence. Both processes are likely to occur – to corroborate this proposition a future study should examine a group of people with a moderate amount of tags.

This study highlights a need for individuation. While the concept of 'folksonomy' heralds the development of cultural capital in the form of commonality of interests, when it comes to describing people, we would recommend the augmentation of uniqueness and areas of specialty. Tagging system designers should consider the deep difference between tagging people and objects. They should not import the same technology used for objects and apply it to people, but they are advised to consider the goals of tagging people, such as impression management, and design the system accordingly.

Finally, tag clouds are designed so that popular tags appear in large and bold font and are more prominent in the tag cloud than less popular tags. Is this a good practice when it comes to people tagging? Do we prefer to identify the maximum number of people based on large, bold, tags or do we want the tagging system to help us to identify the

uniqueness of people, be it on a professional or personal basis? We leave these as questions for further thought and discussion indicating the importance of this study for system design.

ACKNOWLEDGEMENT

This work is partially supported by the EC co-funded project SocIoS. SocIoS is an FP 7 STREP project (Project number: 257774. Call Identifier FP7 ICT 2009.1.2: Internet of Services, Software and Virtualisation). For more information see http://www.sociosproject.eu.

The work on this paper was partially supported by an Open Collaborative Research grant from IBM.

REFERENCES

[1] Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B. and Gosling, S. D. Facebook profiles reflect actual personality, not self-idealization. Psychological Science, 21, 3 (2010), 372.

[2] DiMicco, J., Millen, D. R., Geyer, W., Dugan, C., Brownholtz, B. and Muller, M. Motivations for social networking at work. In Anonymous *Proceedings of the 21st ACM conference on Computer supported cooperative work*. ACM, , 2008, 711-720.

[3] Dominick, J. R. Who Do You Think You Are? Personal Home Pages and Self-Presentation on the World Wide Web. Journalism and Mass Communication Quarterly, 76, 4 (1999), 646-658.

[4] Ehrlich, K. and Cash, D. Turning information into knowledge: Information finding as a collaborative activity. In Anonymous *Proceedings of Digital Libraries.* (1994), 119-125.

[5] Ellison, N., Heino, R. and Gibbs, J. Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment. Journal of Computer-Mediated Communication, 11, 2 (2006), 415-441.

[6] Farrell, S., Lau, T., Nusser, S., Wilcox, E. and Muller, M. Socially augmenting employee profiles with people-tagging. In Anonymous *Proceedings of the 20th annual ACM symposium on User interface software and technology.* (2007) , 91-100. ACM New York, NY, USA.

[7] Gibbs, J. L., Ellison, N. B. and Heino, R. D. Self-Presentation in Online Personals: The Role of Anticipated Future Interaction, Self-Disclosure, and Perceived Success in Internet Dating. Communication Research, 33, 2 (2006), 152.

[8] Goffman, E. *The presentation of self in everyday life.* Allen Lane, , 1959.

[9] Joinson, A. N. European Journal of Social Psychology, 31, 2 (2001), 177-192.

[10] Kaid, L. L. and Wadsworth, A. J. Content analysis. Measurement of communication behavior, (1989), 197–217.

[11] Kenny, D. A. *Interpersonal perception: A social relations analysis.* The Guilford Press, , 1994.

[12] Lindgaard, G., Fernandes, G., Dudek, C. and Brown, J. Attention web designers: You have 50 milliseconds to make a good first impression! Behaviour & information technology, 25, 2 (2006), 115-126.

[13] Marcus, B., Machilek, F. and Schütz, A. Personality in cyberspace: Personal web sites as media for personality expressions and impressions. J. Pers. Soc. Psychol., 90, 6 (2006), 1014.

[14] Muller, M. J., Ehrlich, K. and Farrell, S. *Social tagging and self-tagging for impression management.* TR 06-02. IBM Research technical report, 2006.

[15] Papacharissi, Z. and Rubin, A. M. Predictors of Internet use. Journal of Broadcasting & Electronic Media, 44, 2 (2000), 175-196.

[16] Raban, D. R., Ronen, I. and Guy, I. Acting or Reacting? Preferential Attachment in a People Tagging System. Journal of the American Society for Information Science & Technology, 62, 4 (2011), 738-747.

[17] Raban, D. R. Self-Presentation and the Value of Information in Q&A Web Sites. Journal of the American Society for Information Science and Technology, 60, 12 (2009), 2465-2473.

[18] Rafaeli, S., Raban, D. R. and Kalman, Y. Social Cognition Online. In Amichai-Hamburger, Y. ed. *The social net: The social psychology of the internet.* Oxford University Press, Oxford, England, 2005.

[19] Rouse, S. V. and Haas, H. A. Exploring the accuracies and inaccuracies of personality perception following Internet-mediated communication. Journal of research in personality, 37, 5 (2003), 446-467.

[20] Savicki, V., Kelley, M. and Oesterreich, E. Judgments of gender in computer-mediated communication. Comput. Hum. Behav., 15, 2 (1999), 185-194.

[21] Schau, H. J. and Gilly, M. C. We are what we post? Self-presentation in personal web space. Journal of Consumer Research, (2003), 385-404.

[22] Singletary, M. W. *Mass communication research: Contemporary methods and applications.* Longman White Plains, NY, 1994.

[23] Trammell, K. D. and Keshelashvili, A. Examining the new influencers: A self-presentation study of A-list blogs. Journalism & Mass Communication Quarterly, 82, 4 (2005), 968-982.

[24] Turkle, S. *Life on the Screen: Identity in the Age of the Internet.* Simon & Schuster, New York, 1995.

[25] Vazire, S. Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. J. Pers. Soc. Psychol., 98, 2 (2010), 281.

[26] Walther, J. B. Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. Communication Research, 23, 1 (1996), 3.

[27] Whitty, M. T. Revealing the [] real'me, searching for the [] actual'you: Presentations of self on an internet dating site. Comput. Hum. Behav., 24, 4 (2008), 1707-1723.