Chapter IX

Exploiting Collaborative Tagging Systems to Unveil the User–Experience of Web Contents: An Operative Proposal

A. Malizia

Universidad Carlos III de Madrid, Spain

A. De Angeli

University of Manchester, UK

S. Levialdi

Sapienza University of Rome, Italy

I. Aedo Cuevas

Universidad Carlos III de Madrid, Spain

ABSTRACT

The User Experience (UX) is a crucial factor for designing and enhancing the user satisfaction when interacting with a computational tool or with a system. Thus, measuring the UX can be very effective when designing or updating a Web site. Currently, there are many Web sites that rely on collaborative tagging: such systems allow users to add labels (tags) for categorizing contents. In this chapter the authors present a set of techniques for detecting the user experience through Collaborative Tagging Systems and we present an example on how to apply the approach for a Web site evaluation. This chapter highlights the potential use of collaborative tagging systems for measuring users' satisfaction and

discusses the future implications of this approach as compared to traditional evaluation tools, such as questionnaires, or interviews.

INTRODUCTION

Collaborative tagging is the process by which users add metadata to a community-shared content, in order to organize documents for future navigation, inspection, filtering, or search. The content is organised by descriptive terms (tags), which are chosen informally and personally by the user. The freedom to choose unstructured tags is the main distinctive feature of collaborative tagging systems, as compared to traditional digital libraries or other systems of content organization, where the creation of metadata is the task of dedicated professionals (such as librarians) or derives from additional material supplied by the authors (Bennis et al. 1998, Csikszentmihalyi, 1997). Like all socially-generated structures, tagging is an adaptable process; it takes the form best supported by the content, letting users decide the categorization of such content, rather than imposing a rigid structure on it. Collaborative tagging is most useful in an environment like the World Wide Web, where a single “content classification authority” cannot exist and there is a large amount of data content being continually produced by the users.

The widespread success of collaborative tagging systems over the last few years has generated a large collection of data reflecting opinions on, and evaluation of, web contents. In this chapter, we look into the possibility of exploiting this large database to evaluate the user experience (UX) of web sites. UX is a multi-faceted construct recently introduced into the HCI agenda to describe the quality of an interactive system (Garrett 2003; McCarthy and Wright 2005). This construct is used to indicate how people feel about a product and their pleasure and satisfaction when using it (Hassenzahl and Tracktinsky, 2006). Responses

such as aesthetic judgments, satisfaction or frustration, feelings of ownership and identity are the most prominent aspects of user experiences investigated in this new, comprehensive, HCI research area (De Angeli, Sutcliffe and Hartman, 2005; Hartman, Sutcliffe and De Angeli, 2007; Norman, 2004). Normally, these responses are collected in formal evaluation settings via questionnaires and/or interviews. Collaborative tagging may offer an interesting alternative, one which is cheaper and less prone to experimental bias. In this chapter, we present a technique to extract semantics from tagging systems, and interpret them to describe the user experience when interacting with on-line content.

This chapter has the following organisation. Paragraph 2 reviews related works on collaborative tagging systems. Paragraph 3 describes three different techniques that can be used to extract semantics from tagging systems. Paragraph 4 reports a method to derive semantics differential attributes from collaborative tagging systems, 3, and its evaluation. Paragraph 5 summarizes the chapter, delineates future trends in the use of collaborative tagging systems for automating evaluation techniques and draws the conclusions.

BACKGROUND

Collaborative Tagging Systems (Golder et al., 2006; Mathes, 2004) offer their users the possibility to index contents for organizing web-related information, sharing knowledge and opinions. There is a growing number of successful web sites which include collaborative tagging, allowing users to index and share different types of contents. Del.icio.us (<http://del.icio.us/>), for example, specializes on bookmarking, categoriz-

ing and sharing URLs, Flickr (<http://www.flickr.com/>) allows users to tag photographs they own; Technorati (<http://technorati.com/>) is devoted to tag weblogs; and Youtube (<http://www.youtube.com/>) allows tagging videos. Other interesting examples are Snipit (<http://www.snipit.org/>), which offers the functionality of bookmarking sections of web pages, and CiteULike (<http://www.citeulike.org/>) or Connotea (<http://www.connotea.org/>) that allow tagging and commenting references to academic publications.

Collaborative tagging systems allow users to become active contributors in the classification of web-content. Because of this characteristic some authors refer to them as “folksonomy” (Mathes, 2004), short for “folk taxonomy”, albeit there is still some debate whether this term is accurate (Golder et al., 2006). Users of collaborative tagging systems do not only categorize information for themselves, but they can also share their classification and browse the information categorized by other users. In fact, many collaborative tagging systems have features for sharing contents and their associated tags among users. They also, offer functionalities for keeping contents private, shared only within a pre-set list of users, or public (shared with everyone). Therefore, tagging is both a personal and a social activity. According to the number of people who can tag the same content and/or to the level of privacy of the tag (shared vs. personal). Collaborative tagging systems are distinguished in “broad” and “narrow” systems (Van der Wal, 2005). A broad tagging system is the result of one item being categorized by many people (Del.icio.us is an example). This can generate a very diverse set of tagging, as different users can enter their preferred terms, with obvious semantic and syntactic variations. There will be some terms that are used by many people to describe one item or many items which are described by the same terms. The concentration of terms can take advantage of power laws (like the Zipf distribution (Zipf, 1949; Newman, 2005)) to quickly see the preferred terms for an

item or items.). It states that the frequency of the occurrence of a term is inversely proportional to its frequency class. Zipf has discovered experimentally that the more frequently a word is used, the less meaning it carries.

A narrow folksonomy is the result of one person categorizing one item (Flickr is an example). In this case, tags are private, but users could decide to share their own photos allowing others to view their tags and thus their categorization of contents. When the contents (and tags) are shared with other users a narrow folkosomy can approximate a broad one; nevertheless since the option of sharing contents and tags is leaved to the final user we cannot strictly rely on it.

This paper concentrates on broad collaborative tagging systems, where several users index and share different content. We regard the folksonomy produced by these systems as a result of collective intelligence and social creativity (Fischer 2006): different users contribute to the establishment and dissemination of knowledge. In this vision, collaborative tagging systems are not only important for their primary task (e.g., information retrieval), but they assume a fundamental role in the quest for understanding the user experience. People tag content with words which have both denotative and connotative meaning; these tags are a reflection of their opinion on the content, the service provider and the interface design. We believe that tagging systems act as social dynamics enablers representing the real “vox populi”; in fact, users can take advantage of tagging information shared by others (Nov, 2007). Tagging systems leave the users free to express their own opinion without restricting them in a frame, such as a questionnaire. We believe that this method is more likely to capture the ecological perception of the web site audience. Collaborative tagging systems offer a lot of unstructured metadata (tags) associated to many different contents (web sites, photos, videos, etc.) that can be used for measuring the UX of these contents over the Internet.

Moreover, collaborative tagging systems allow detecting variations over time, by analyzing how tagging evolve. The goal of our research is to develop a methodology to extract meanings from collaborative tagging systems and to use this information in order to understand what people think about on-line contents. This methodology requires a two phased process: (a) detecting semantics from tagging systems; (b) interpreting the meaning of this information.

DETECTING SEMANTICS FROM TAGGING SYSTEMS

Information retrieval (IR¹) from unstructured contents such as those produced by tagging systems) is a complex task. A major problem relates to the fact that no current tagging systems have synonyms control (e.g. “Mac” and “Macintosh” do not coincide in Del.icio.us). For this reason, in order to use the information contained in a collaborative tagging system, we need to use techniques extracting semantics from users’ tags. In the following paragraph, we discuss three information retrieval techniques that can be used to extract semantic features from tagging.

Many information systems use keywords or key phrases to search or browse collection of documents for specific terms and information. Not only are keywords used for searching relevant documents but also to index and categorize the content. Relevant information is indicated by the authors of a document and is placed in appropriate sections for emphasizing them. Typical examples are the title, abstract and author’s name written with bold or in appropriate places of the document. This approach is useful if employed within document collection explicitly managing this information, such as newspapers articles. Nevertheless this information is not available in general and providing them manually can be tedious or inapplicable depending on the amount

of relevant keywords or terms we want to provide for each document.

IR algorithms were devised to address this problem, trying to automatically extract relevant terms and keywords from unstructured document collection. IR algorithms employ two different phases (Turney, 2002): keywords assignment, and keywords extraction. Usually, there is a training phase where an initial default list of relevant keywords is provided to the system, thus using a controlled dictionary. The wider is the list, the greater should be the number of documents used to train the system by manually indicating the keywords included in each document (chosen among the given list). These types of algorithms are called training-intensive, i.e. a big training set is required to obtain good performance.; On the contrary, keywords extraction does not need any training since the keywords are directly extracted from the body of each document by using the information learned from the training phase and some similarity measure. In the next section, we present a selection of three IR algorithms that can be used to automatically extracting semantics from collaborative tagging systems.

PMI-IR

The PMI-IR (Point wise Mutual Information – Information Retrieval) algorithm employs the technology of a search engine, such as Google Page Rank, or Yahoo, (Krikos, et al. 2005; Kraft et al. 2006)) to extract the frequency of searched keywords within a collection of documents. In general, the algorithm takes as input a word and a set of alternative terms for that specific word. The output is the selection of the terms whose meaning is the closest to the given word. That is to say, the algorithm finds the synonyms by analyzing the co-occurrences of the terms with the given keyword and among them.

This is exactly the case for tagging systems, where we have a collection of contents labeled

with different words representing keywords for that collection and we would like to group words having the same meaning.

There exist different ways of measuring the co-occurrence between two terms, but the one used by PMI-IR algorithm is based on the Point wise Mutual Information (1), where problem represents the given word (tag in a folksonomy), $\{choice_1, \dots, choice_n\}$ represent the n alternatives for problem and $P(\text{problem}, choice_i)$, $i=1, \dots, n$ the probability of the co-occurrence.

$$\text{Score}(choice_i) = \log_2 (P(\text{problem}, choice_i) / P(\text{problem}) P(choice_i)) \quad (1)$$

If problem and $choice_i$ are statistically independent, then the probability of co-occurrence is described by $P(\text{problem}) P(choice_i)$. If problem and $choice_i$ are not independent (i.e., they tend to co-occur) the numerator in (1) will be greater than the denominator and the ratio will describe the independence rank between the two terms.

By considering that $P(\text{problem})$ is assuming the same value for each associated $choice_i$ and that the log function is monotonically increasing, equation (1) can be simplified as follows:

$$\text{Score}(choice_i) = \log_2 (P(\text{problem}, choice_i) / P(choice_i)) \quad (2)$$

The conditional probability value $P(\text{problem} | choice_i)$ is assigned as a measure of how close the words are (synonyms). This measure can be computed, for instance, by using a search engine like Google page rank or Altavista advanced search. $P(\text{problem} | choice_i)$ represents the number of documents returned by the search engine, called hits, when searching for problem and $choice_i$. The term which is the most similar to the problem is the one that maximizes the measure as shown in (3).

To clarify how this algorithm can be used for extracting semantics from tags, let us give an example. In the first instance we consider as

co-occurring two words appearing in the same document, for example tags used in del.icio.us for categorizing the same web site; e.g. www.micro-soft.com tagged with both the words 'explorer' and 'windows'. In this context, the score assigned to each $choice_i$ is computed as follows:

$$\text{Score}(choice_i) = (\text{hits}(\text{problem AND } choice_i) / \text{hits}(choice_i)) \quad (3)$$

The equation reported in (3) assigns as score the value of the ratio between the number of documents containing the two terms (problem and $choice_i$) and the number of documents containing only $choice_i$.

The tag which is most correlated to the problem is the one obtaining the highest score value computed as of (3). This is a reasonable is a reasonable measure of similarity among tags and a given term, yet it has some problems. In fact, a good similarity measure should include the totality of the tags included in a tagging system and not only within a basic set of $choice_i$ terms. Different tags can have different meanings depending on the interpretation of the author of the tags. Thus, this approach is suitable for narrow folksonomies where clusters of tags are created by some users who upload the contents using their own way of categorizing contents which are likely to have a narrow and well defined range of synonyms.

Collaborative Tag Suggestion

A new IR algorithm (Xu et al., 2006) has been recently introduced, which is based on tag suggestions for annotating documents in collaborative tagging systems. This method assigns reputation weights to the authors of tags, on the basis of the accuracy of words entered. The system suggests terms to use as tags for documents based on the words which are most frequently used by users with good reputation (good sense-making).

This algorithm still keeps into account the magnitude of the co-occurrences of terms but using a subset of terms used by certain experienced users. The objective is to evaluate which tags are relevant (keywords) for the documents in the folksonomy. This objective is achieved by a ranking among users indicating which ones participate positively (the most reliable) on the tagging process. The notation used for this algorithm is defined as follows:

- $P_s(t_i/t_j; o)$: it is the probability that a user labels an object o with the tag t_i knowing that the tag t_j has already been used for the same object (i.e., document). To measure the correlation between the two tags on the object o , the algorithm considers the ratio between the number of users using both t_i and t_j , and the number of users using only t_j .
- $P_a(t_i/t_j)$: it is the probability that any object is labeled with the tag t_i knowing that the tag t_j has already been used for the same object. In this case the observation refers only to the tags and not to the objects. To measure this correlation between the two tags, the algorithm considers the ratio between the number of users using both t_i and t_j , and the number of users using only t_j .
- $S(t, o)$: indicates the score of the tag t on the object o , computed by summing the number of users that labeled o with t .
- $C(t)$: indicates the coverage of a tag, which is the number of objects labeled with t . The greater is the number of objects tagged with t , the less specific is the meaning of the tag t . In other words if t is used very often it is a generic term.

The algorithm works by iterating the selection of the tags t_i for which $S(t, o)$ is high and multiplying this by the inverse of $C(t_i)$. After selecting the t_i with the maximum score, the scores of

every other tag t' are changed according to the following statements:

- t' score $S(t', o)$, is decreased removing redundant information, i.e. subtracting the value of the probabilities product of t' and t_i used together. In this way the superposition of the suggested tags is reduced, as in $S(t', o) = S(t', o) - P_a(t'/t_j) S(t_j, o)$
- t' score $S(t', o)$, is increased if it co-occurs with the selected tag t_j over the object o , as in $S(t', o) = S(t', o) + P_s(t'/t_j; o) S(t_j, o)$. This procedure allows dealing with basic level variations of tags, normalising the score for tags like BLOG, BLOGGING, and BLOGS.

The drawbacks of this approach are related to the fact that there exist narrow folksonomies, like Flickr, where every object o (i.e. a document) is own by the user that uploaded it, or by users to whom the owner has granted access permissions. In these cases, we do not have access to all the information needed for running the algorithm over a wide number of tags. Thus the clustering process does not necessarily represent the users' opinion on the tagged topic; but a specific feeling about the shared content (e.g. a link to a web site, and thus the web site itself) can be detected by means of finding related words among a community of users sharing the same interests.

The Semantic Halo algorithm

In our previous work we introduced a Semantic Halo technique in order to deal with word semantics in tagging systems (Dix et al., 2006). The basic idea consisted of using co-occurrences of tags to cluster their relationships and meanings.

The Semantic Halo is defined as a set of search results for a given tag made by a set of four features, labeled as **4A**:

- **Aggregation.** Representing all the tags linked or related to a given tag.
- **Abstraction.** It is similar to aggregation but it relates to a direction (increasing and decreasing), thus it contains two subsets:
 - o Generalization, tags increasing abstraction with respect to the given tag,
 - o Specialization, tags decreasing abstraction with respect to the given tag.
- **Ambience.** It is the context for a given tag. It includes all the possible tags appearing in the same context, and that will be useful for augmenting or refining the user query. This set is built from a basic context set.
- **Age.** It is a list of the Ambience feature elements over time. It helps in retrieving tags ordered by meanings given to them over time.

The algorithm was tested within the Del.icio.us community. In this environment, users submit their links to a website, adding some descriptive text and keywords, and Del.icio.us aggregates their posts with everyone else's submissions allowing users to share their contributions. The algorithm was implemented using Del.icio.us programming APIs (Application Programming Interfaces). This procedure allowed to collect results while the users were tagging. Because the Del.icio.us community is very large and active this test resulted in a quite complex but effective test.

For example, given the tag "university", which is quite general, our algorithm searched over Delicious for related tags and retrieved:

Ambience = {'open', {'learning', 'University'}}
Abstraction = {'online', 'education'} U {'colleges', 'high', 'degree', 'distance', 'Commons'}
Age = (('learning', 'University'), ('open'))²
Aggregation = {'soccer', 'gradschool', 'corps', 'indoor', 'course', 'masters', 'research in-

stitute', 'cites', 'cincinnati', 'peace', 'demographic', 'content', 'courses', 'innovators', 'urban', 'tournament', 'entrepreneurship', 'liverpool', 'york', 'community-college', 'schools', 'Illinois', 'abroad', 'Content', 'latino', 'Course', 'complexity', 'planning', 'Initiative', 'academiclibrary', 'enterprise', 'semantic-web', 'Education', 'grad', 'scholarship', 'teaching', 'college', 'school'}.

We can observe that the Ambience set is composed of two subsets, associated with two different contexts or meanings of the 'university' tag. The algorithm can solve also basic level variations³ since the tag 'University' with the capital 'U' is strongly associated with the 'university' tag (without using any parser). The first part of the Abstraction set is related to generalization of the given tag, while the second part is specialization, thus providing a partition of the related tags in increasing and decreasing abstraction. The Age sequence is the ordered set of contexts (meanings) with respect to last updates. The Aggregation set lists all the related tags, and even if there are unwanted tags the majority (as shown in the example above) is clearly related.

The Semantic Halo algorithm is applicable in general to broad and narrow collaborative tagging systems but has the drawback of employing a clustering technique that can be less effective or precise in specific sub-domains originated by users' tags.

Summary

All the different algorithms presented in these paragraphs can be employed for extracting semantics from tags by automatically organising them in classes or synonyms. The designer can choose different algorithms or techniques depending on the characteristics of the considered tagging system.

PMI-IR is quite fast to compute and since it is a standard approach within the IR field, many

implementations can be found. The drawbacks of this approach are related to the fact that tags have different meanings depending on the sense-making of the users. As a consequence, for retrieving useful semantics the algorithm should span over the entire collection of tagged contents (considering the different choice_i words). The Collaborative Tag Suggestion approach is very effective but it should be avoided when dealing with narrow tagging systems (Flickr for example) when every tagged content is owned by the user or shared with a specific subset of users granting permissions to them. Finally, the Semantic Halo can be used to extract semantics both from broad and narrow Collaborative Tagging Systems but it is less precise in specific sub-domains of tags (users' annotating contents in on specific domain or topic).

Choosing an IR algorithm is a first step for organizing unstructured content, which is a prerequisite for evaluating the UX over contents shared by tagging systems.

SEMANTIC DIFFERENTIAL IN COLLABORATIVE TAGGING SYSTEMS

This section describes a method to evaluate the information extracted from the tags in order to obtain a measure of the user-experience with web-sites. The evaluation phase in our approach is based on the elaboration of the concept of semantic differential introduced by the psychologist and communication scholar E. Osgood (1975). The original work of Osgood focused on the measurement of meaning, addressing issues of word semantics and psychological differences between words. In his influential research the author proposed a method (the semantic differential) to highlight individual differences in the attribution of meaning to words. The semantic differential measures people's reaction to stimulus words and concepts. Participants are invited to rate the

stimulus with a bipolar scale. Each extreme of the scale is labeled by contrasting adjectives, such as bad-good. This technique has been frequently used in psychometrics to measure a number of psychological constructs, and more recently has been employed in HCI to build user satisfaction questionnaires. An example of opposite couples of adjectives used by Osgood methodology is shown in table 1.

Osgood research has demonstrated that ratings on bipolar adjective scales tend to be correlated, and to cluster around three basic dimensions of response, which account for most of the co-variation in ratings. These dimensions, labeled as Evaluation, Potency, and Activity (EPA), have been verified by factor analyses and replicated in an impressive variety of studies.

In our approach, there are no fixed couples of opposite adjectives but the information is extracted from the adjectives freely introduced by the user. The adjectives are then associated to one of the three dimensions evaluation, potency, and activity

Table 1. Opposite couples of adjectives used by Osgood

1.	Angular/Rounded,
2.	Weak/Strong,
3.	Rough/Smooth,
4.	Active/Passive,
5.	Small/Large,
6.	Cold/Hot,
7.	Good/Bad,
8.	Tense/Relaxed,
9.	Wet/Dry,
10.	Fresh/Stale.

Evaluation

In order to test our methodology of semantic differential through collaborative tagging systems we built a basic collaborative tagging system. Users could add tags for categorizing contents, we explicitly asked users to use adjective as tags for categorizing the contents. Then by employing an IR algorithm we obtained a structured representation of the tags. Successively, we clustered structured tags into groups of adjectives. These groups of adjectives were used for applying the semantic differential technique and obtaining a measure of the UX on the web-site contents.

The objective of our evaluation study was to evaluate the experience of a community of users with respect to the Sapienza University of Rome Italian web portal (www.uniroma1.it). We choose this target because many users of the web portal were complaining about its features and usability. In fact, the Sapienza web portal has now been redesigned.

The community considered in this experiment is composed of 48 people. The majority of users (60%) were students. The remaining sample was split in 20% of administrative staff and 20% of academic staff.

Participants were invited to browse the web-site and tag it with their preferred set of adjectives. This system provided users with classic collaborative tagging functionalities, such as: presenting the

document to be tagged and the text labels where the corresponding tags could be added. Figure 2 shows the tagging systems used for the experiment, even if displayed in Italian the tagging and web site area are clearly visible.

We collected around 162 tags from 48 individuals in a 2 weeks time-frame. As a first step we analyzed the frequency of each tag (adjective) as presented in table 2. Analysing the adjectives, it appears that there is a sort of binary distribution of the general tags among positive and negative evaluation. Looking at the frequency distribution (for example $f=11$), we found that two very different tags (simple and dazed) are the most

Table 2. Table showing the most occurring tags

Tag	Occurrence
Simple	11
Dazed	11
Clear	6
Sad	6
Comprehensive	6
Inconsistent	5
Intuitive	4
Useless	4
Poor	4

Table 3. Classes assigned after the clustering phases and number of tags falling in their relative class; the third rows represent the number of tags per class with positive and negative meanings, (+n) stands for n tags in that category with positive meaning and (-n) for the negatives.

Evaluation	Potency	Activity
82 (+40)(-42)	52 (+11)(-41)	28 (+8)(-20)

frequently used. This effect is evident for almost every couple of tags in Table 2. The effect was also evident in the complete dataset, even with less frequent tags. This let us hypothesize that even after the clustering process (assigning tags to the three classes: evaluation, potency and activity) the user perception would be split in two neat categories according to the overall binary perception: positive or negative.

We categorized the tags in the three classes according to the clustering proposed by Osgood in (Osgood et al, 1975), as shown in Table 3. We used the PMI-IR to automatically measure the distance between the selected adjective (Problem) and the couples of adjectives contained in the

Osgood scale. This approach has been employed for the positive and the negative meanings of a tag. The three major factors for a tag to belong to a class are:

- Evaluation: representing the overall feelings about the web site (adjectives like good or bad);
- Potency: representing the expressive power and impact on the perception of the web site (adjectives like strong or weak);
- Activity: representing the possibilities and functionalities (at informational level) offered by the web site (adjectives like: active or passive).

Figure 1. The custom collaborative tagging system including tagging labels and interested web site

Figure 2. A bar-chart view of the users with respect to the selected classes (evaluation, potency and activity)

Figure 3. A bar-chart view of the overall evaluation, potency and activity

By using the clustering results we analyzed the experience of each user, expressed by the tags inserted in the system with respect to the selected classes (evaluation, potency and activity) and their positive or negative meaning. Figure 2 shows individual results for each user (1-47⁴). The three colours represent the classes, while each sector displays the normalised score and category to which each tag inserted by the users belongs to (radar view).

Figure 3 suggests that, in general, the user evaluation of the web-site has been quite negative as most of the scores fell in the negative half of the scale. Potency is the weakest dimension. Furthermore this graph highlights that users with a positive evaluation of the web site focus their attention to that particular class (the Evaluation class), which deals mainly with strong feelings about a web site (adjectives like: good or bad, nice or ugly, etc.).

CONCLUSION

This chapter presented a UX evaluation approach consisting of three steps: 1) select a collaborative tagging system containing the content for which we would like to evaluate the users' experience (Delicio.us for web sites, Flickr and YouTube for multimedia content, Technorati for blogs); 2) employ an IR technique to extract semantics for users' tags and group them together by cluster of synonyms or related tags; 3) use the data automatically extracted from the clusters of tags to detect the overall impression of users (UX) over the selected content (web sites, multimedia systems, blogs, etc.) by grouping tags according to the semantic differential technique. We reported an example of how this evaluation approach can be applied on an ad-hoc collaborative tagging system. Anyway, this procedure can be applied to a wide range of collaborative tagging systems. In our test we asked users to add adjective tags

to keep it controlled but generally we will have plenty of available tags already inserted by users of collaborative tagging systems. What do you mean by that? You need to explain further by re-writing this sentence.

Our approach suggests the importance of collaborative tagging systems in the evaluation of the end users experience. It seems to be a promising and cost effective alternative to questionnaires or interviews.

Collaborative tagging systems are becoming increasingly popular on the Internet. There are many reasons why users are motivated in volunteering their time to support these on-line communities (Clary et al., 1998), adding information to collaborative tagging systems over the web; nevertheless such systems keep growing as a social phenomena. We can take advantage of this huge number of users to detect the user experience perceived by them when adding tags for categorizing a content of interest over the web.

REFERENCES

- Bennis, W., & Ward Biederman, P. (1998). None of us is as smart as all of us. *IEEE Computer*, 31(3), 116–117.
- Clary, E. G., Snyder, M., Ridge, R. D., Copeland, J., Stukas, A. A., Haugen, J., & Miene, P. (1998). Understanding and assessing the motivation of volunteers: A functional approach. *Journal of Personality and Social Psychology*, 74(6), 1516–1530.
- Coleman, W. D., & Williges, R. C. (1985). Collecting detailed user evaluations of software interfaces. *Proc. Human Factors Society: Twenty-Ninth Annual Meeting*, Santa Monica, CA, (pp. 204-244).
- Csikszentmihalyi, M. (1998). Creativity: Flow and the psychology of discovery and invention. *Perennial*, June 1997.

- De Angeli, A., Sutcliffe, A., Hartmann, J. (2006). *Interaction, usability and aesthetics: What influences users' preferences? DIS 2006 Conference Proceedings, ACM*, (pp. 271-280).
- Dix A., Levialdi S., & Malizia A. (2006). Semantic halo for collaboration tagging systems. In S. Weibelzahl & A. Cristea (Eds.), *Proceedings of Workshop held at the Fourth International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH2006). Workshop on Social Navigation and Community-Based Adaptation Technologies* (pp. 514-521), Lecture Notes in Learning and Teaching, Dublin: National College of Ireland.
- Fischer, G. (2006). *Distributed intelligence: extending the power of the unaided, individual human mind*. In C. Augusto (Ed.), *AVI* (pp. 7-14). ACM Press.
- Garrett, J. J. (2003). *The elements of the user experience: User centred design for the web*. London: Easy Riders..
- Golder, S., & Huberman, B. A. (2006). Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32(2), 198-208.
- Hartmann J., Sutcliffe A., & De Angeli, A. (2007). Assessing the Attractiveness of Interactive Systems. *CHI 2007 Conference Proceedings* (pp. 387-396), San Jose, CA, USA,.
- Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *Behaviour & Information Technology*, 25, 91-97.
- Kraft, R., Chang, C. C., Maghoul, F., & Kumar, R. (2006). Searching with context. In *Proceedings of the 15th international Conference on World Wide Web* (pp. 477-486) (Edinburgh, Scotland, May 23 - 26, 2006). WWW '06. ACM, New York, NY.
- Krikos, V., Stamou, S., Kokosis, P., Ntoulas, A., & Christodoulakis, D. (2005). DirectoryRank: ordering pages in web directories. In *Proceedings of the 7th Annual ACM international Workshop on Web information and Data Management* (pp. 17-22) (Bremen, Germany, November 04 - 04, 2005). WIDM '05. ACM, New York, NY.
- Mathes, A. (2004). *Folksonomies cooperative classification and communication through shared metadata*. Retrieved October 12, 2006, from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>.
- McCarthy, J., & Wright, P. (2005). *Technology as Experience*. Cambridge: MIT Press.
- Newman, N. (2005). Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46, 323-351.
- Norman, D. (2004). *Emotional Design: Why we love or hate everyday things*. New York: Basic Books.
- Nov, O. (2007). What motivates Wikipedians? *Commun. ACM* 50, 11 (Nov. 2007), 60-64.
- Osgood, H. M., William, C. E., & Murray, S. M. (1975). *Cross-cultural universals of affective meaning*. University of Illinois Press, Urbana.
- Turney, P. (2002). *Mining the Web for Lexical Knowledge to Improve Keyphrase extraction: Learning from Labeled and Unlabeled Data*, NRC/ERB-1096, July 19, NRC 44947.
- Van der Wal, T. (2005). *Explaining and showing broad and narrow folksonomies*, Personal InfoCloud: February 2005 Archives. Retrieved November 22, 2007, from <http://www.personalinfocloud.com/2005/02/>.
- Xu, Z., Fu, Y., Mao, J., & Su, D. (2006). Towards the semantic web: Collaborative tag suggestions. In *proceedings of the Collaborative Web Tagging Workshop at WWW2006*, Edinburgh, Scotland, May, 2006.
- Zipf, G. K. (1949). *Human behaviour and the principle of least effort*. Cambridge MA: Addison-Wesley

KEY TERMS

Collaborative Tagging Systems: Collaborative tagging (also known as folksonomy, social classification, social indexing and other names) is the practice and method of collaboratively creating and managing tags to annotate and categorize content⁵.

Distributed Intelligence: In many traditional approaches, human cognition has been seen as existing solely “inside” a person’s head, and studies on cognition have often disregarded the physical and social surroundings in which cognition takes place. Distributed intelligence provides an effective theoretical framework for understanding what humans can achieve and how artifacts, tools, and socio-technical environments can be designed and evaluated to empower human beings and to change tasks⁶.

Information Retrieval: Information retrieval (IR) is the science of searching for information in documents, searching for documents themselves, searching for metadata which describe documents, or searching within databases, whether relational stand-alone databases or hypertextually-networked databases such as the World Wide Web⁷.

Semantic Clustering: Identifying and disambiguating between the senses of a semantically ambiguous word, without being given any prior information about these senses⁸.

Semantics Differential: A type of a rating scale designed to measure the connotative meaning of objects, events, and concepts⁹.

Usability Evaluation: Usability usually refers to the elegance and clarity with which the interaction with a computer program or a web site is designed¹⁰.

User Experience: User experience, often abbreviated UX, is a term used to describe the

overall experience and satisfaction a user has when using a product or system¹¹.

ENDNOTES

- ¹ Searching a body of information for objects that match a search query, particularly a text or other unstructured forms (<http://www.cs.cornell.edu/wya/DigLib/MS1999/glossary.html>).
- ² Round parenthesis are used in the mathematical sense that we are not enumerating a set here but we consider an ordered sequence in the case of Age feature.
- ³ Basic level variations are considered to occur when, having two words differing by the case or including or not a dash.
- ⁴ One user has been deleted from the sample because inserted tags as spam, due to the anonymous login to the system.
- ⁵ Collaborative Tagging Systems. (2007, Nov. 20). In *Wikipedia, The Free Encyclopedia*. Retrieved Nov 20, 2007, from <http://en.wikipedia.org/wiki/>.
- ⁶ Fischer, G. (2006). *Distributed intelligence: extending the power of the unaided, individual human mind.*, AVI (Augusto Celentano, ed.), ACM Press, 2006, pp. 7–14.
- ⁷ Information Retrieval. (2007, Nov. 20). In *Wikipedia, The Free Encyclopedia*. Retrieved Nov 20, 2007, from <http://en.wikipedia.org/wiki/>.
- ⁸ Semantic Clustering (2007, Nov. 20). In *Wikipedia, The Free Encyclopedia*. Retrieved Nov 20, 2007, from <http://en.wikipedia.org/wiki/>.
- ⁹ Semantics Differential. (2007, Nov. 20). In *Wikipedia, The Free Encyclopedia*. Retrieved Nov 20, 2007, from <http://en.wikipedia.org/wiki/>.
- ¹⁰ Usability Evaluation. (2007, Nov. 20). In *Wikipedia, The Free Encyclopedia*.

- Retrieved Nov 20, 2007, from <http://en.wikipedia.org/wiki/>.
- ¹¹ User Experience. (2007, Nov. 20). In *Wikipedia, The Free Encyclopedia*. Retrieved Nov 20, 2007, from <http://en.wikipedia.org/wiki/>.