TagReel: A Visualization of Tag Relations among User Interests in the Social Tagging System

Joohee Bae
Integrated Design Laboratory, Department of Media
Graduate School of Ajou University
Suwon, Korea
jooaya0211@ajou.ac.kr

Kyungwon Lee
Division of Digital Media
Ajou University
Suwon, Korea
kwlee@ajou.ac.kr

Abstract— **Social tagging systems provide users with the ability to share information and extend their field of knowledge. The purpose of this paper is to explore the tag relations of user interest in these systems and examine the semantic relations of tag usage in terms of user interests. To do this, a classifying method was used to characterize words into seven classes. Experiments were limited to the five areas of interests: music, photography, games, books and videos. Samples were taken from 50 English-language web pages in the social bookmarking service Delicious. The top 7 tags were extracted from each web page that was tagged as music, photography, games, books and videos. Via the results of the tag analysis of this study, relationship between tags and user interests were demonstrated. TagReel employed a radial visualization method, which has the advantage of giving an overview of the whole cluster of connection among tags. TagReel can represent the semantic concept of connection between tags and between interests. It can contribute to raising the cognitive power and awareness of tag usage in various categories and show the relative weight of tags.**

Keywords-visualization; social tagging; tag; resource; user; social bookmarking

I. INTRODUCTION

Social tagging has received a considerable amount of interest in recent years due to its advantages in sharing information. Through this system, users create a collection of items of personal interest and assign individual keywords to each of the resources in that collection [1].

Individual keywords, referred to as tags, relate to different types of content or to the work of a user. In contrast to categorization, tagging represents a new approach to organizing information. Nonhierarchical classification allows data to be freely gathered, allows easy access, and has the ability to move directly to other types of content. A study of tagging by Riddle (2005) found that tagging helps to increase personal recall and that it has beneficial social effects while allowing for serendipity and novelty [2]. Tagging can freely represent a user's own ideas as keywords and can allow users to manage information individually. Additionally, tagging groups related URLs together [3] and the same tags also cross over diverse fields of resources [4]. Thus, it has the potential to connect a variety of types of content. Users are used to describing resources of interest by means of one or more freely chosen keywords and exploring every user and tags with the purpose of discovering various resources of interest [5]. Accordingly, tags are considered as user interests and collecting a new resource also thought as user activities of interests.

The first part of this paper investigates the structure of social tagging system and illustrates relationships among users, tags, and resources. After ascertaining the major patterns, the second section identifies the classification of tags and conducts a qualitative study of the distribution of tag types. In the last section, the visualization of tags in the social tagging system is demonstrated and summarizes the overall discussion.

II. RELATED WORK

A. Social Tagging

Social tagging refers to the tags people use to collect the various keywords to organize a subject. Tagging makes it possible to annotate freely various resources such as images, videos and web pages. Although there is no rule to making words, users become used to sharing and organizing in an unauthorized way in the social tagging system. The importance of social tagging is that it is a reliable way to perceive participatory behavior. This is because users tag proper words to the documents even when no one is looking, for the purpose of sharing with other users and for their own retrieval.

Weinberger (2005) insists that tagging causes people who belong to the social community to form their interests and to develop a similarity of view [6]. Thus, if users have the same tags, it represents that they are sharing deep commonalities with each other. The purpose of tagging is to manage their own data as well as to share and understand other users' thoughts and ideas.

Therefore, tagging is regarded as a method of communication. For these reasons, websites based on the social tagging system, such as Flicker, Delicious, Last.fm and YouTube, have advantages for observing personal interests as well as public interests, simultaneously.

B. *The Structure of the Social Tagging System*

Social tagging systems have three fundamental elements: users, tags and resources [7]. The details are shown in Table 1.

TABLE I. ELEMENTS OF THE SOCIAL TAGGING SYSTEM

Elements	Define	Example value
Users	People who are tagging on the system	ID
Tags	Keywords that freely annotate various contents	Subjects, document types, tasks, opinions, purposes
Resources	The subjects being tagged	URLs, texts, videos, photos

Social tagging breaks the conventional indexing that uses one-on-one relationships, but instead provides information on multiple relationships through tags [4].

It lets users discover new contents, and it also connects the new users with others who have the same tags. Fig. 1 shows the relationship among users, tags, and resources. These three elements are banded together, making it possible to easily navigate the related topics or new contents.

Figure 1. Relationship among users, tags.

III. METHOD

A. *Classification of Tags*

To analyze characterized patterns of the existing tags, classification of tags were investigated for the present study. Golder and Huberman (2006) initially investigated tag categorization concerning functional scheme and he found a number of tagging patterns [8].

This study expands Golder's concept to include the relationship between the resources and users' thoughts. The new classification of the present study focuses on the characteristics of web pages tagged by Delicious users; therefore, this classification is applicable to the social tagging service Delicious.

The proposed classification of tags is comprised of seven classes, presented as shown below.

- **Subject and topic**: Tags that describe the main service offered by a website such as camera, tutorial and magazine.
- **Type of a website**: Tags that are connected with types of a website, for example, blog, community and flash.
- **Title of a website**: Tags are identical to a site title or name. These are related to the name of company or URLs, for instance, Nintendo, Yahoo and Stock.
- **Facet of a page**: Tags that emphasize the characteristic point of view and the information that the user needs, such as reviews, history and wallpaper.
- **Type of content**: Tags that are connected with types of contents such as images, audio, midi and video.
- **Impression**: Tags that express users' opinion, emotion and feeling when visiting the site. For example, nature, urban and technology.
- **Task**: Tags for the purpose of using the collected sites in terms of personal activities.

B. *The Data*

For this study, samples were obtained from Delicious, a good example of the social tagging system. Delicious is one of the well-known sites of the social bookmarking service. This site has the ability to manage a list of bookmarks on a webpage and to share that list. In addition, it is possible to add a list created by other users to an original list. Thus, it easily catches the users' interest, spontaneously. To discover the relationship among interests, the datasets were limited to five interests: music, photography, games, books and videos. The representative tags were then determined for each interest. The representative tags were from the popular tag of each category. Via popular tags, a considerable amount of tags are allowed to be gathered and have a bias towards access to the more frequently used ones.

Resources that are known as bookmarks were collected from April, 2009. The number of URLs was randomly examined. They were 50 English-language web pages that were tagged with music, photography, games, books and videos, respectively. Samples are extracted from the top 7 tags per each webpage. The analysis of a tag frequency from resources conducted by Kipp and Campbell (2006) found that the top 7 tags were important parts to describe a specific site. In his study, consensus determined the top 7 tags as most representative of the document [3]. The extracted samples are described in Table 2 in brief.

The number of tags from Table 2 excluded duplicate tags. For example, if users tag a web page twice with the same tags, it would only count as one tag. Therefore, those samples of tags are unique.

TABLE II. EXTRACTED SAMPLES FROM DELICIOUS

Category	Representative tag	Number of tags
Music	music	206
Photography	photography	179
Game	games	205
Book	books	215
Video	video	205

C. *Method of a Tag Analysis*

In this section, distribution of tags was examined through seven classes of tag classification to discover the relationship among tags that reflected user interests. The method in this study was classifying the data from Delicious into seven classes by considering the characteristics of the resources. The method was to reduce the error that is caused by an equivocal meaning; the author visited the websites and searched the terms. That was a laborious work and needed a lot of care. Tags are subjective, with the consequence that the method of a qualitative study was conducted rather than a quantitative.

The method of a qualitative study was decided on to identify what kinds of tags are clustered around a specific type and to compare distribution of seven classes in the five areas of interests. Although tags are same, it would be possible to understand the differences by means of characteristics of categories. For example, 'news' is considered a subject and topic in the area of music, however, in the case of games, it may be thought of as facets of a page.

D. *Distribution of Tag Types*

Here are some results as follows.

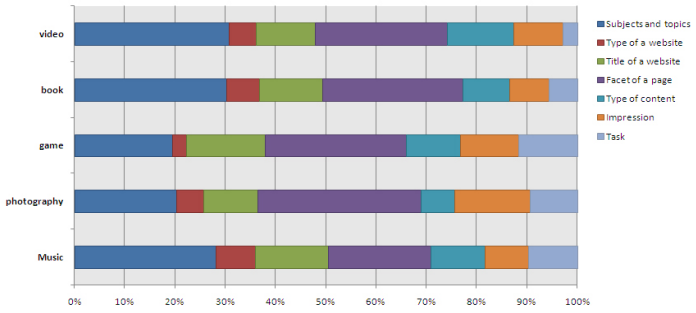

Figure 2. Distribution of seven classes

Fig. 2 represents the tag variation of the five areas of interests in seven classes. Those samples of tags are concentrated comparatively on the classes of a subject and topic, and facet of a page. However, there are some differences in each category which depend on the characteristics of music, photography, games, book and video. For example, music tends to concentrate on the subject and topic class. And lots of tags are exposed on the facets of a webpage in the area of photography. In the case of the area of game tags are exposed on the facets of a page higher than other classes as well as a type of a website class is the lowest in the seven classes. Both book and video have a great deal of tags in the two classes of a subject and topic, and facet of a page. Furthermore, task class is the lowest part in the area of video. Table 3 shows a list of example tags from a result.

TABLE III. EXAMPLE OF TAGS IN SEVEN CLASSES

Seven Classes	Music	Photography	Game	Book	Video
Subject and topic	radio, reviews, records, download	camera, tutorial, guide, portfolio,	emulator, abandonware	bookcovers, library, magazine	Advertising, converter,
Type of a website	blog,mp3blog, community	blog, flash, community	blog, flash	blog, catalog, wiki	screencast, community
Title of a website	pandora, mtv, youtube, emusic	flicker, timelapse, candid, stock	nintendo, brain, phonics, bitorrent	pelican, mercurial, martinfowler	balloon, graphisme, voki
Facet of a page	streaming, free, discography	free, howto, sharing, history	freeware, p2p, crowdsourcing	encyclopedia, literature, bookmarking	History, parody, tracking
Type of content	audio, mp3, video, sound, midi	photos, images, digital, graphics	flash, music	fantasy, pdf, textbooks	audio, documentary, film
Impression	vintage, aggregator, culture	nature, urban, inspiration	fun, interactive, humor, funny,	charity, functional, spirituality	inspiration, trends, cool
Task	storage, shopping, player	retouch, tools, reference	reading, shopping, benchmark	reference, research, tools	traning, ideas, tools

IV. VISUALIZATION

From the results of the tag analysis of this study, relationships between tags and user interests were demonstrated. TagReel - the visualization of this study - represents the distribution of Delicious tags in seven classes, and shows the differences of a tag usage within the five areas of interests: music, photography, games, books and videos.

A. An overview and comparison of visualizing tag relations

TagReel tracks a variety of global flows in a circular interface. First, it tries to let the users compare the variety of distribution of seven classes in five areas of interests. Second, it shows the differences of the same tags' usage in different categories. Third, frequencies of tags are represented to distinguish how or what the popular tags are, intuitively.

Fig. 3 is a result of the visualization of tags. This visualization is able to trace the same tags among the five areas of music, photography, games, books and videos and provides insight into key routes within each interest.

Words at the edge of the circle represent unique tags that were extracted from Delicious. The dots beside the words indicate the frequency of the tags.

Five dots are the maximum, and that represents that the same tags are exposed more than 16 times in a class. It is possible to intuitively compare the distribution of tag frequencies in this one picture.

Starting at the 3 o'clock position in a clockwise direction, the five areas of interests are separated by space: photography, music, videos, games and books.

Color bars in the inner part of the circle represent the tag distribution of seven classes. Starting at the red color in a clockwise direction, these seven classes are situated thusly: subjects and topics, types of a website, titles of a website, facets of a page, types of content, impressions and tasks. The length of each color bar indicates the proportion of the number of tags in a specific class.

Curved lines in the circle show the connection between tags. The color of the relation curves, connecting blue to blue or red to red, demonstrates linking between tags in the same class as well as in the other interest areas. For example, it may connect 'download' (of the subject and topics class) in the category of music to 'download' (of the facet of a page class) in the category of game. The gradient colors of the relation curves, connecting blue to red or orange to green, show the linking between tags in the different classes as well as in the different interest areas. They show the divergences of tag usage among the five categories with connecting the same words.

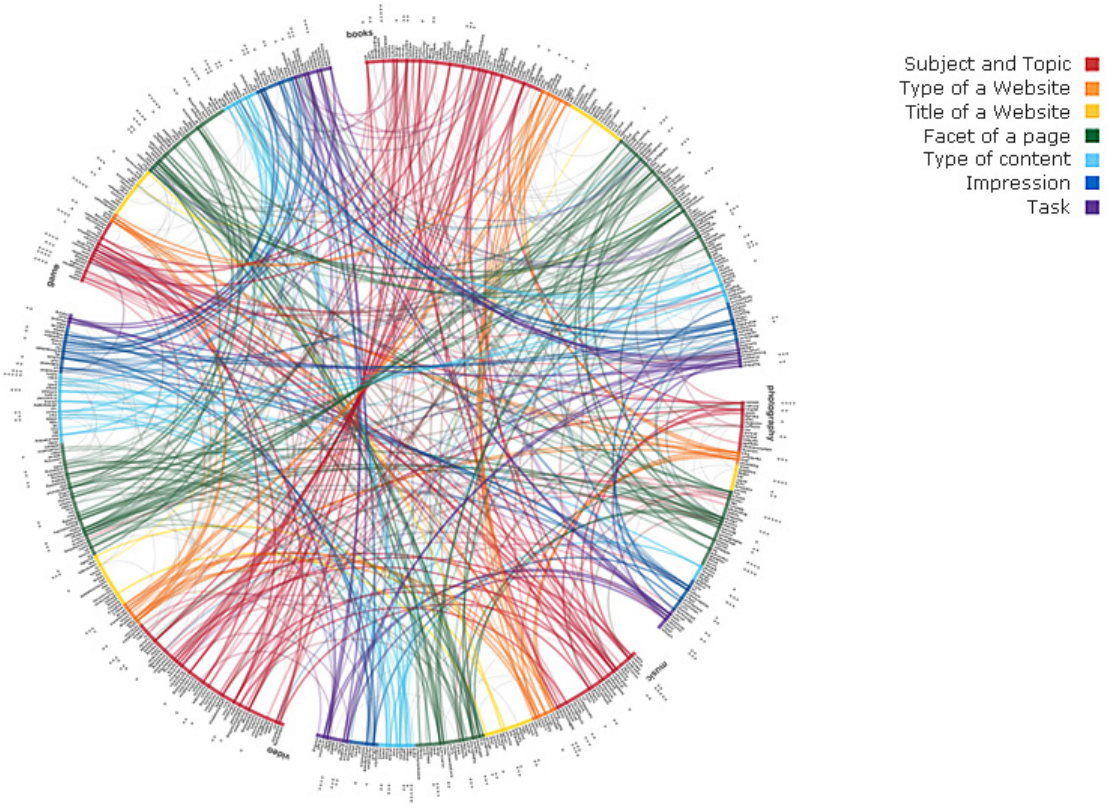

Figure 3. TagReel: Visualization of tag relations in the five areas of interests

A curved line of gray illustrates the linking between tags in the same category as well as in the different classes. For example, 'review' belongs with the class of facets of a page in the area of game; however, it is also included in a subject and topic class. These relation curves demonstrate that tags have a various meanings, even if they are in the same category.

B. *Radial methods*

The concept of radial has been widely applied in visualization research. Radial visualization [9] is a newly coined term by Hoffman et al. (1997). It arranges data in an elliptical way and has visual effects on a variety of subject areas. A recent study, conducted by Draper et al. (2009), identified seven design patterns of radial visualization [10]. The present study draws on the idea of a Connected Ring-based visualization scheme, one of the seven design patterns. The connected ring pattern [10] places nodes around the edge of the ring and connects nodes with lines. If additional nodes need to appear, those are optionally located in the ring's interior.

This has the advantage of giving an overview of the whole pattern of connections by viewing relationships among miscellaneous collections.

C. *Navigations*

The visualization of this study allows simple navigation just by clicking operations.

1) Selection: Selection is for the purpose of comparing two or more areas of categories. When users click the left button of the mouse on a name of the five categories, selected clusters will be highlighted. When users click it one more time, it will be unselected.

2) Focus + Context: Fous + Context allows users to show detailed information linked with the context, and it is also beneficial to focus on other information while users are interacting [11]. Focus + Context works under the premise that users want to see both overview and detail view simultaneously, and these two types of views can be combined within a single display using a dynamic interaction [12]. Fig. 6 is an example of the detail view.

D. *Visualizing techniques*

Fig. 3 was developed in Action Script 3.0 with the XML (eXtensible Markup Language) data.

1) Curves: The curves were made as parts of circles, known as 'Arcs', for aesthetic purposes. To make the Arcs, the Bezier Curve algorithm was used. Fig. 4 is a part of the programming language for drawing an Arc.

```
// Draw curved lines
arc.drawArc(sCenterX, sCenterY, sRadius,
            angleTwoPts*-1*180/Math.PI, startAngle*180/Math.PI);
```

Figure 4. Action script for drawing an Arc

2) Gradient color: Gradient is represented by measuring angles between two nodes. The changing point of color is at the center of a curved line. Fig. 5 shows a detailed examination of this.

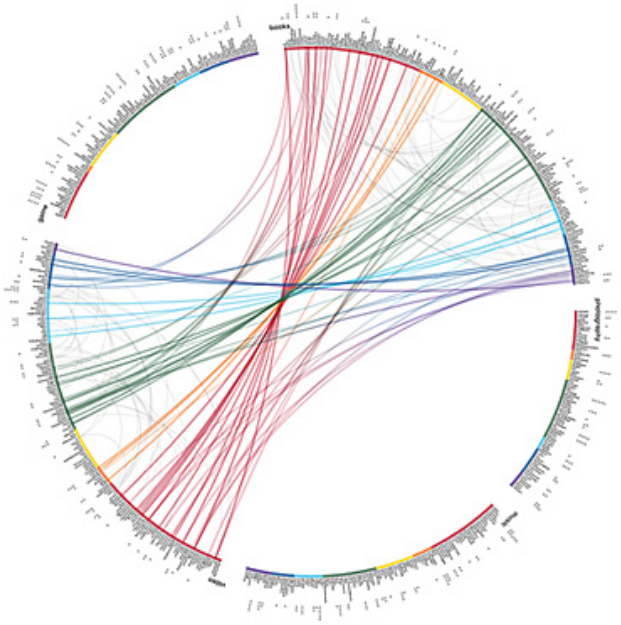

Figure 5. Comparison the two areas of books and videos

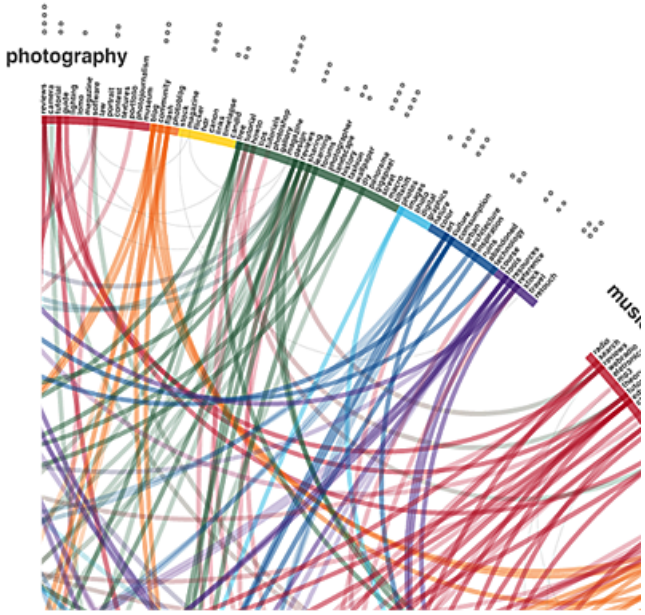

Figure 6. Detail View

V. CONCLUSIONS

This paper has dealt with the tag relations of user interests in the Social Tagging System. First, this paper determined the relationships among users, tags and resources in social tagging systems and examined the semantic relations of tag usage in terms of user interests. Second, to analyze the characteristics of tags when adding a new web page to a list, experiments were conducted. Samples were from the social bookmarking service Delicious. They were taken from 50 English-language web pages that were tagged as music, photography, games, books and video, respectively. Those samples of tags were then classified into seven classes by considering the characteristics of the resources. Third, via the result, relationships between tags and user interests were demonstrated. The radial method was used, and tags were located around the edge of the ring and connected with lines. This method easily showed the relationships among miscellaneous collections and gave an overview of the whole pattern of connections

TagReel has the ability to compare the variety of distribution of seven classes in the five areas of interests and to show the differences of the same tags' usage in different categories while tracing the same tags. It also gives a perspective in which users apply a tag to a webpage and establish the capacity of expanding a social service that offers the opportunity to create a new business model.

ACKNOWLEDGMENT

We would like to specially thank Kunhwan Park, who is an undergraduate student in Ajou University, for implementing the application in our study.

REFERENCES

[1] Markus Heckner, Tanja Neubauer, and Christian Wolff, "Tree, funny, to_read, google: what are tags supposed to achieve?," In Proceedings of the 2008 ACM Workshop on Search in Social Media (SSM 08), ACM, October 2008, pp. 3-10, doi:10.1145/1458583.1458589.

[2] Prentiss Riddle, "Tags: what are they good for?," The University of Texas at Austin, December 2005, pp. 1-12.

[3] Margaret E.I. Kipp and D. Grant Campbell, "Patterns and inconsistencies in collaborative tagging systems: an examination of tagging practices," In Proceedings of of the 2006 Annual Meeting of the American Society for Information Science and Technology (ASIST 06), November 2006.

[4] Ruey-Ming Chao, Sheng-Wen Lo, and Jia-Nan Chang, "Using tags to construct the digital resource management system," Ninth IEEE International Symposium on Multimedia (ISM 07), IEEE, 2007, pp. 464-469, doi:10.1109/ISM.Workshops.2007.84.

[5] Francesco Ronzano, Andrea Marchetti, and Maurizio Tesconi, "Tagpedia: a semantic reference to describe and search for web resources," Workshop on Social Web and Knowledge Management (SWKM08), ACM, April 2008.

[6] David Weinberger, "Taxonomies to tags: from trees to piles of leaves," Release 1.0, CNET Networks, February 2005.

[7] Harry Halpin, Valentin Robu, and Hana Shepherd, "The complex dynamics of collaborative tagging," In Proceedings of the 16th international conference on World Wide Web (WWW 07), ACM, May 2007, pp. 211-220, doi:10.1145/1242572.1242602.

[8] Scott A. Golder and Bernardo A. Huberman, "Usage patterns of collaborative tagging systems," Journal of Information Science, vol. 32, April 2006, pp. 198-208, doi:10.1177/0165551506062337.

[9] Patrick Hoffman, Georges Grinstein, Kenneth Marx, Ivo Grosse, and Eugene Stanley, "DNA visual and analytic data mining," In Proceedings of the 8th IEEE Visualization '97 Conference, IEEE, October 1997, pp. 437-441, doi:10.1109/VISUAL.1997.663916.

[10] Geoffrey M. Draper, Yarden Livnat, and Richard F. Riesenfeld, "A survey of radial methods for information visualization," IEEE Transactions on Visualization and Computer Graphic, in press.

[11] Juan C. Dursteler, "Focus + Context," The digital magazine of InfoVis.net, April 2002.

[12] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, "Readings in information visualization: using vision to think," Morgan Kaufmann Publishers, 1999.