Method of Evaluating Contents on the Basis of Community's Interest Using Data from Social Bookmark Services*

Y, Fukami
Graduate School of Media and Governance, Keio University
5322 Endo Fujisawa-city, Kanagawa, Japan
yofukami@sfc.keio.ac.jp

T, Sekiya†
Faculty of Engineering, The University of Tokyo
7-3-1 Hongo Bunkyo-word, Tokyo, Japan
t50861@mail.ecc.u-tokyo.ac.jp

I, Ohmukai and H, Takeda
National Institute of Informatics
2-1-2 Hitotsubashi Chiyoda-word, Tokyo, Japan
{i2k|takeda}@nii.ac.jp

Abstract

This research aims toward the development of multifarious information search technique by utilizing contents which are accumulated to SBM and tags. With the time series analysis of tagging conduct, the method can extract contents which are appraised as sprouting of SBM users' interest, as central interest of them, and as the topic of communication between them respectively. In other words, tags of SBM can bring us appropriate information according to our intent. Since tags tell us major interest and cutting edge of the users', the way extracting from tag has potential working for profiling users and making use of the information for marketing activity toward them. We conduct approach to make use of actual data to search for polyphenic situation which has never been conducted. We think this way must be improved when you recognize mechanism of SBM as communication platform and the function of tags.

Keywords: SBM, tag, time series variation, new tag ratio, search

1 BACKGROUND

1.1 INFORMATION OVERLOAD ON THE INTERNET

Recently, due to the popularity of full-time internet connection, the increase of networking speed, expansion of storage capacity, CPU processing ability and decreasing prices makes the amount of data being stored or transferred on the internet to increase (Ministry of Internal Affairs and Communications, Japan, 2006). On the other hand, the amount of information a person can process have its limit as what have been mention by Simon (1996) in past times and, recently Davenport and Beck (2001). From a user's perspective, it's necessary for individuals to accurately select the data or information that are effective to them. However, paralleling to the increase in data, there emerged fluctuation in data quality. In fact, it is becoming more difficult for us to find necessary information from the internet.

*The data for the analysis for this thesis was provided by EC Navi Company. The authors would like to express their deepest gratitude.

†Present contact address: Hakuhodo Incorporated

1.2 INTRODUCTION OF INFORMATION ORGANIZATION TOOLS

When gathering information on the internet, most frequently used tool are the search engines. In Japan, there are some book guiding effective way to find appropriate information using search engines, especially about advanced search technique using Google: c.f. Tsuda (2004) and other titles. Like this, many efforts to effectively use the search engines can be seen. However, search engine shows only a list of the sites and ranking that are relevant to the query. The user needs to access all the listed sites and locate relevant contents to find the contents meeting his/her own need. Even if the user can bookmark and save the URL of interesting sites, there are limits on the amount of bookmarks. It will be also difficult to find information from the overpopulated bookmarks.

Several technologies, like weblog and social networking service (SNS) have been developed to simplify the process of information exchange over the internet. Much of consumer's voice and experience has been distilled on these kind of sites, then they are collectively called as CGM (Consumer Generated Media). Collecting and organizing massive information on the internet is a time consuming process. On the other hand, recently new tools have been developed and introduce to ease this process. One of it is the social bookmarking service (SBM).

1.3 CHARACTERISTICS OF THE SOCIAL BOOKMARKING SERVICE (SBM)

Social bookmarking service (SBM) is a service where a website's URL can be saved on the server. Basically, it is an implementation of bookmarking function available on the browser over at the server side. The following is what differs from the normal browser based bookmarking functions:

1. When saving the URL, can provide annotation such as keywords or comments.
2. The URL list and annotation would be publicly accessible.

However, if the URL was just saved only, utilizing this massive accumulated URL later would be difficult. The bookmarking functions on the browser basically managed by separating data into folders. On the other hand, SBM, instead of using folders to manage the data, metadata that were saved as keywords or comments can be used as keys during search process. Additionally, the accumulated URL and metadata are publicly available, and possible for other users to use it.

1.4 CLASSIFICATION BY FOLDER AND TAGGING

To organize files and data on a computer, we create folders and classify by directories. When classifying by folders, understanding the hierarchy will be a premise. If the hierarchy was not understood properly, it is difficult to search (Marlow et al., 2006) and if multiple folders were created, it will be difficult to locate the folder (Jones et al., 2005).

Recent years, the tagging classification method have been introduced and widely used on web services. The idea of tagging is adding some words (mostly one word as a keyword) as a metadata to the targeted data or URL. The words used for tagging are called tag. When a tag was issued, a link automatically formed with the respective list. By tagging, the respective file to be saved will be grouped and classified. Tags are being used in a variety of areas for classification purpose.

For example, Technorati[1] uses tags to classify weblog, and Gmail[2] uses tags to classify emails. With Technorati, it classifies individual's writes blog articles. While with Gmail, it classifies e-mails sent and received. With SBM, you can tag websites (URL) and the tagged information can be shared with other users. In other words, SBM is a massive database for websites with metadata.

[1] Technorati:http://technorati.com/

[2] Email service provided by Google Inc (URL: https://mail.google.com)

2 DESIGN TO ANALYSIS

2.1 USER COMMUNITY FORMATIONS

With SBM, the accumulated data is publicly available. The history of all user's bookmarking and tagging are calculated frequency to display at ranking page. Individual's collection of accumulated sites reflects the individual's preferences. Data aggregated in this service which consists of URL and tags, is a reflection of the user's preferences. SBM was designed to use other people's accumulated information, and we assume that users with similar preferences can come together, forming a community with similar preferences.

2.2 VOCABULARY USED ON TAGS AND RECOMMENDED INFORMATION BY TAGS

On the SBM the vocabulary used for tagging represent the classification of the users. The collections of tags are a reflection of their preferences. It is an indication and visualization of the users' information management system. Collection of the tags is a representation of all the users' information management system by preferences. Based on this concept, "folksonomy" a classification system from bottom up emerged (Mathes, 2004). The ideas to aggressively utilize tag groups or classification system as folksonomy were established as information distribution system.

Niwa et al. (2006) proposed the approach to recommend information with using accumulated tags and relation between tags and users. First, gather information from whole system of the SBM, and calculate affinity between vocabulary used for the tag and user respectively. Then cluster vocabulary on the basis of affinity to avoid fluctuation. Lastly, calculate the sense of affinity of each keyword cluster to each user, and then recommend user bundle of sites scaled with growth of attention. On the other hand, Ohmukai et al. (2006) proposed collaborative recommendation method on the basis the classification system not of network allover, but between users with close relationship, to bring more precise information according to individual's preference.

2.3 VOCABULARY CONCENTRATION

There were multiple proposals on information search supporting method, as mentioned. However, there were problems with using tags as a medium for sharing information. The vocabulary used on tags tend to be concentrated (popular vocabularies) to small numbers, and it could be hard to reach information suited variety of perspectives. For this analysis, we used data from SBM service, `Buzzurl`[3] operated by EC Navi Company. The top 1.1% of the popular vocabularies used covers 50% of all the tags (Figure.1). It is too concentrated in small number of vocabulary to make use of tags for the query of searching because to lead disproportionate recommendation.

2.4 CHANGE IN THE TIME SERIES OF THE VOCABULARY BY TAGGING

To prevent these concerns, we looked at the frequency used, and also the change in time series of the vocabulary used. There is a research to visualize the time series change of the vocabulary used (Dubinko et al., 2006). In contrast, we did not only follow the change, but also checked the new vocabulary used. We framed a hypothesis that users adopt new vocabulary for tag when the tagged contents contain something new aspect for them, and we emphasized on these new vocabulary used.

Tags are used to organize the accumulated contents. Therefore, users tend to use similar tags to annotate similar contents. Users must not increase the variation of tags more than necessary. New tags are used on contents that differ from ever accumulated sites, or on contents accumulated from a different perspective. We want to apply this thinking to all the users. By implementing this, sites many users annotate with new tags must contain information many users think are new.

[3]Buzzurl:http://buzzurl.jp/. The service name was EC Navi Popular News when data was provided

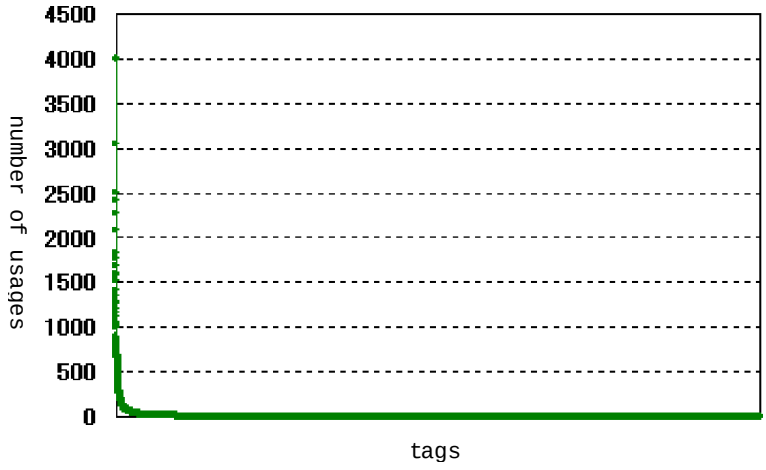

Figure 1: Frequency distribution of vocabulary for tags

2.5 NEW TAG RATIO

Newly used vocabularies by the user are named as new tags. Whether the vocabulary is already used or not by anyone, we determine individual tag as new tag or not, according to the each user's history of usage. In other words, if other users have ever used the vocabulary, and even if the user has never adopted it before, we will treat it as a new tag (Figure.2).

user name	post date	tags	site title
userA	5-May	design	NextDesign Leadership Institute
userA	6-May	design	swfIR: swf Image Replacement
userA	7-May	design	AnyGivenFriday - Home
userA	8-May	design	Made With Molecules
userA	9-May	design	InfoQ: How to Design a Good API & Why it Matters
userA	10-May	design	45 Fresh, Clean and Impressive Designs | Smashing Magazine
userA	11-May	design	jQuery App-Like menu with ul's/li's
userA	12-May	design	Web Developer's Handbook: developing web-sites, exploring imagination
userA	13-May	food	Hot Fudge Pudding Cake

Vocabulary that was newly used = new tag

no	user name	post date	tags	site title
1	userA	5-May	design	NextDesign Leadership Institute
2	userA	6-May	design	swfIR: swf Image Replacement
3	userB	7-May	food	The Oil We Eat (Harpers.org)
4	userB	8-May	food	CREAMY CHICKEN-NOODLE CASSEROLE
5	userC	9-May	food	National Center for Home Food Preservation
6	userC	10-May	food	All recipes – complete resource for recipes and cooking tips
7	userB	11-May	food	Hot Fudge Pudding Cake
8	userC	12-May	food	Hot Fudge Pudding Cake
9	userA	13-May	food	Hot Fudge Pudding Cake

User B/C already used the tag "food", No. 7/8, will not considered as a new tag.
User A, used "food" at No.9, for the first time, hence it will become a new tag.

Figure 2: Definition of new tag.

We classified stored sites as total number of whole tags issued, and ratio of new tags issued by each URL. We name the ratio as *new tag ratio*. Therefore, new tag ratio is indicated by formula below (1), c: contents (site), R: new tag ratio, n: the total new tags, and t: all the tags.

$$R(c) = \frac{\sum_n(c)}{\sum_t(c)} \tag{1}$$

We consider site catching the attention of many users can be bookmarked by many users, and contents that have high new tag ratio include new information many users are interested in.

3 ANALYSIS SUBJECT

3.1 SELECTING SBM ANALYSIS SUBJECT

Based on the assumption that we have showed up to now, we will initiate a verification analysis of actual stored data on SBM: accumulated URLs and tags.

We used `Buzzurl`(Figure.3)[4], for analysis, as stated previously. The concept of this service, similar to digg[5] is to register news related site, and extract contents with new information. This is suitable for our intent for the analysis. Buzzurl is operated by a Japanese corporation called EC Navi Company. This site is in Japanese and majority of the users, are using Japanese.

Figure 3: Buzzurl

3.2 CHARACTERISTICS OF BUZZURL (FORMERLY EC NAVI POPULAR NEWS)

3.2.1 FUNCTION TO GENERATE FURTHER COMMUNICATION

Buzzurl does not only specialize in genre of contents, but also provide users functions which characterize the site suitable for handling news contents. The first user to register a new sites is displayed largely (Figure.4), and this promote other users to register new sites.

They provide a service to promote communication between users, too. Users can register someone with similar interests or gathering many contents you valuate good as a *favorite user*. The list of sites registered by Favorite Users is displayed in one's individual page. Also, the site displays the users who tend to register similar sites to one's accumulation as *Synchro-user*. Users can simply find other users with similar interests and preferences. Buzzurl generates further communication among users with using these functions.

3.2.2 TAGGING SUPPORT

When registering a site, candidates for tag are displayed. These consist of a list of vocabulary used on similar sites or one's own history of usage (Figure.5). These tagging supports lead users to select vocabularies ever used by themselves or other users.

[4]Buzzurl:http://buzzurl.jp/

[5]digg:http://www.digg.com/SBM, specialize in news related contents, providing services in English

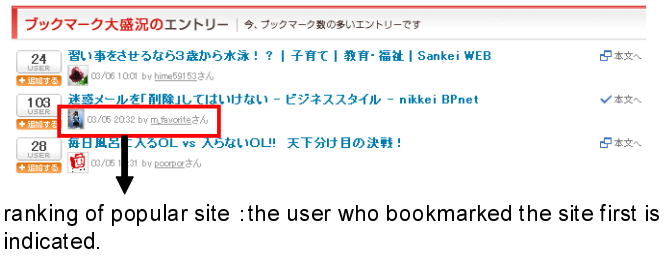

Figure 4: Site ranking of the highest registered number: user that registers first on the respective site will be displayed

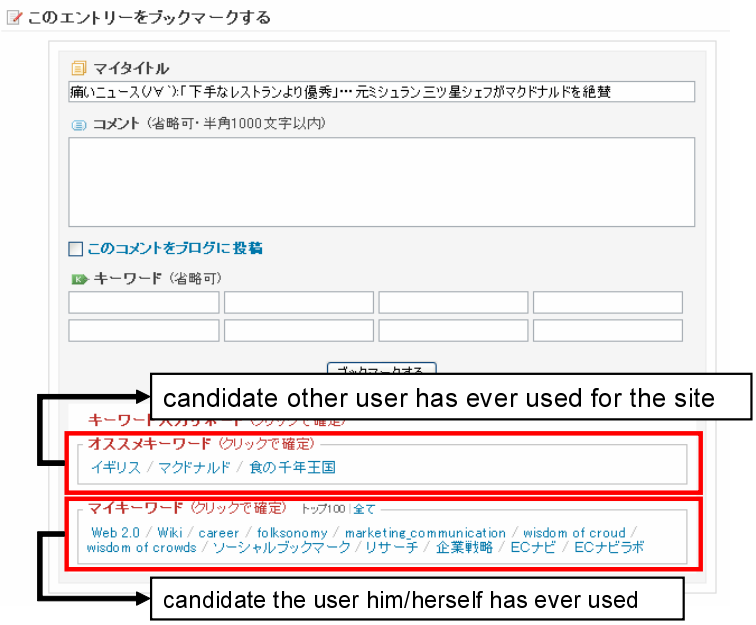

Figure 5: Tagging support

3.3 SUMMARY OF THE ANALYZED DATA

Data used for analysis is the registered bookmark data from October 2005 to December 2006. During this time frame, users registered to this site were 1,221 users. There were 74,128 types of registered URL, and registered number was 139,602. Out of 139,602 bookmarks, there were 264,058 tags. 31,076 types of vocabularies used, per 1 bookmark the average number of tags were 1.9.

3.4 USER CHARACTERISTICS

The type of users that registered the most number of URLs were professionals of planning & marketing (11.7%), researchers, developers and technicians (10.8%), followed by engineers (9.9%). Top 3 categories of business registered over 30% of the total users. From the user population perspective, the dominant were the housewives[6] (18.6%), however they did not use the SBM functions aggressively. The core users were those professionals involved with planning and marketing

[6]The reason for having many housewives was that EC Navi Company, which operates Buzzurl, originally was a price comparison site. Furthermore, it was due to the fact that the members from the price comparison site were issued a Buzzurl account without processing a new membership registration.

or engineering.

occupation	bookmarked sites(%)	number of users(%)	average number of bookmarking
all	16382sites	130586users	-
planning, marketing	11.7	1.7	56.3
technical expert	10.8	9.1	9.4
engineer	9.9	3.9	20.4
part time job	8.8	9.0	7.8
homemaker	7.8	18.6	3.4
student	7.6	5.4	11.3
PR, advertising, design	4.7	1.1	33.7
sales person	4.6	8.9	4.1
self-employed individual	3.4	5.6	4.8
company executive	3.2	1.9	13.5
managerial worker	2.3	2.3	7.8
clerical pos	2.0	5.0	3.2
general secretary, personnel worker	1.4	2.7	4.1
civil officer	0.5	2.8	1.5
accountant	0.5	2.2	1.8
specialist personnel	0.4	2.1	1.5
fringe worker	0.4	2.9	1.1
teacher	0.1	0.9	0.7
Agriculture, Forestry and Fisheries	0.0	0.2	0.7
Others	20.1	13.6	11.7

Table 1: number of sites bookmarked by Buzzurl user by occupation

4 TAG ANALYSIS

4.1 URL REGISTRATION DISTRIBUTION

The frequency distribution of the registered URL is shown in Figure.1 before, a small amount of sites were registered intensively and a large number of sites were registered by small number of users, representing a long tail distribution.

On the tags of SBM, there were problems with "polysemy," "synonymy", "basic level variation"(Marlow et al., 2006). Yet for this analysis, if strings differed, we treated as a different vocabulary. For example, "EC navi" and "EC Navi", were counted as separate vocabulary.

4.2 TREND OF THE VOCABULARY USED FOR THE TAGS

When analyzing a large number of tags that are being used, "Google", "Web2.0", "YouTube", "Microsoft", computer or internet related vocabularies were frequently used in most cases. Also, "advertising" came in 14th, responding to the characteristics of the people who are actively registering the sites. In other words, shows the interests of people that belong to planning & marketing and engineering business categories.

4.3 SITE DISTRIBUTION BY NUMBER OF REGISTRATION AND NEW TAG RATIO

The average tag per site was 8.4, and among that the amount of tags with new vocabulary, were 3.0. In other words, the new tag ratio was 33.1%. Next, we looked at the relationship of the total amount of tags per site and the amount of new tags. Figure.6 shows a comparison of total amount of tags per site to the ratio of new tags.

rank	tag	number of usage	%	%(accumulation)
1	google	4,004	1.52	1.52
2	neta (troll)	3,041	1.15	2.67
3	funny	2,493	0.94	3.61
4	professional wrestling	2,419	0.92	4.53
5	graphical content	2,278	0.86	5.39
6	SNS	2,086	0.79	6.18
7	Blog	1,826	0.69	7.54
8	showbiz	1,773	0.67	7.54
9	football	1,685	0.64	8.18
10	FIFA World Cup.	1,603	0.61	8.79
11	music	1,594	0.60	9.39
12	web2.0	1,580	0.60	9.99
13	youtube	1,513	0.57	10.56
14	advertising	1,426	0.54	11.10
15	game	1,355	0.51	11.62
16	motion picture	1,344	0.51	12.13
17	social	1,282	0.49	12.61
18	news(in Japanese)	1,262	0.48	13.09
19	sports	1,199	0.45	13.54
20	EC navi	1,164	0.44	13.98
21	news(literal)	1,131	0.43	14.41
22	Microsoft	1,115	0.42	14.83
23	mobile phone	1,053	0.40	15.23
24	movie	1,050	0.40	15.63
25	security	1,011	0.38	16.01
26	mixi(SNS in Japan)	1,002	0.38	16.39
27	convenient	897	0.34	16.73
28	search	888	0.34	17.07
29	gourmet	880	0.33	17.40
30	professional baseball	856	0.32	17.73

Table 2: Top 30 vocabularies used for tags

Using the total amount of tags and new tag ratio on the site, we would like to extract the main interest and the new concerns of the users. Regarding the extraction, the avobe standard was applied based on the average and standard deviation for the total tags and new tag ratio (Table.3).

(1) Sites with many total tags, and little amount of new tags:
the standard was a total amount of tags, over 20, and the new tag ratio, less than 20%.

(2) Sites with many total tags, and new tags:
the standard was total amount of tags, over 20, and the new tag ratio, over 50%.

(3) Sites with little amount of total tag, and many new tags:
the total amount of tags, less than 8, and new tag ratio, over 50%.

According to each standard, we will look at the top 20 registered sites and the top 20 frequently used vocabulary on the tags.

5 RESULTS

5.1 SITES WITH MANY TOTAL TAGS AND LITTLE AMOUNT OF NEW TAGS

First, we will look at (1)sites with many total tags, and little amount of new tags. On Table.4, is listed the top 20 sites in total tags which the total was over 20 and new tag ratio was less than

	number of all tags annotated	number of tags with new vocabulary	new tag ratio (%)
Average	8.42	2.97	33.1
SD	9.20	7.34	27.3

Table 3: average and standard deviation: number of annotated tags

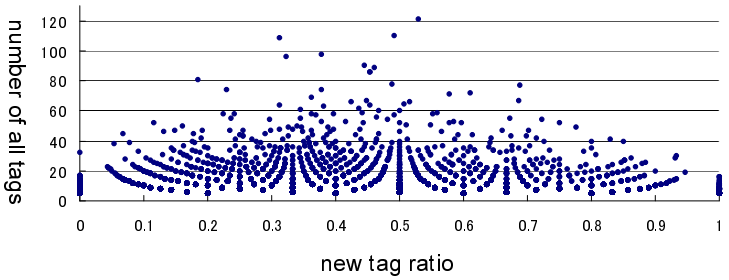

Figure 6: number of tags and new tag ratio

20%. Most sites within the top 20 were ITC or marketing related topics. We can assume that it covered the contents of interest of the active users.

Then, what kinds of vocabularies for tagging on these sites are used? On Table.5 the top 20 most used vocabularies as tags are listed. When compared to the top tags of the overall registered URL(Table.2), the majority of the registered with low new tag ratio has 2 characteristics.

First, it is mainly covered by ITC or marketing related vocabularies and other contents are few. While there are vocabularies as "neta (troll)", "funny", "professional wrestling", "graphical content" in the list of the top 30 sites overall, you cannot find these words in the list of this segment. In the top 5, internet related vocabularies such as "Google", "YouTube", "SNS", "Web2.0", and "Yahoo!" were listed.

Second, the top vocabularies used hold a high ratio on tagging. Of the overall tags, the vocabularies that covered the top 30 had a ratio of 18%, a little short of 20%. Sites on registered new tag ratio that are low, of the top 20, it was 31%, and exceeded 30%. Therefore, it can be said that sites with many tags and low new tag ratio, and the tags annotated to them reflect the core of the active user's interest and concerns.

5.2 SITES WITH MANY TOTAL TAGS, AND MANY NEW TAGS

Next, we will look at (2)sites with many total tags, and many new tags. The contents of the top 20 sites (Table.6) were mainly ITC related, however there were contents like "taste ranking of favorite Umai Bo (Traditional Japanese molded confection)", "What is your favorite Miyazaki movie?", "Doraemon Finale, read in comic", including contents from daily life or culture. Then, what kinds of vocabularies were used on the tags for these sites?

On Table.7, the top 20 vocabulary most frequently tagged is listed, and the number 1 was the operating company "EC Navi(in Japanese)". When compared to Table.6, EC Navi related titles, it shows the concern and interests in respect of the operating company. In other words, loyalties are starting to form for companies that provide SBM platform and this service. From number 2, vocabularies like "neta (troll)" "funny", "news", "ranking", are listed. These words are used to have fun for the moment, or in order to share a topic with others, therefore it cannot be considered as users prior concerns. In other words, these tags were a medium for communication between users, rather than individual's interest and preferences.

Additionally, when site and contents of the tags are considered, a platform called SBM, forms loyalty, which promotes information accumulation. Furthermore, SBM through sites and tags is considered to be a medium of communication. Henceforth, we believe that there is a need to

rank	title of site	number of all tags	new tag ratio(%)
1	Google pledges $900 million for MySpace honors	81	18.52
2	Japan.internet.com column/ search engine marketing	52	11.54
3	ITmedia News : mixi (SNS in Japan) and YouTube affect websphere in Japan	50	16.00
4	Google Moves to Sell Space for Video Spots on Network of Web Sites–New York Times reported - CNET Japan	47	14.89
5	NTT DoCoMo cooperate with Google to provide mobile search- CNET Japan	47	19.15
6	misconception about amazon.com and "long tail" - guide to new economy on the Internet - nikkeibp.jp	46	13.04
7	Ameba (blog service in Japan) redesign of services - CNET Japan	45	6.67
8	Any: VideoSNS for mobile phone increase of capital - CNET Japan	45	17.78
9	ITmedia News : banner ad in mixi (Japanese SNS), number of impressed over times second to Yahoo! Japan	42	19.05
10	Japan.internet.com Web marketing - Can alliance of Yahoo! and eBay threat Google ?	40	17.50
11	Buzzurl has new feature - CNET Japan	39	7.69
12	Any: Japanese SNS start to compatible for video - CNET Japan	39	17.95
13	4 requirement of Web2.0 site - GIGAZINE	38	5.26
14	ITmedia News : Microsoft report to kick around taking over eBay	37	16.22
15	ITmedia D mobile : mobile phone featuring Gachapin (Japanese TV character) released	36	19.44
16	media and PR (blog site): digg and YouTube lounch new service promised to success	35	14.29
17	Using "Google"(verb) in Japanese, OK? - CNET Japan	34	17.65
18	EC navi company launch SBM compatible for video - CNET Venture View	34	14.71
19	way to access Buzzurl via mobile phone | What is Buzzurl? (blog site)	34	14.71
20	Google's antisocial downside - CNET Japan	33	18.18

Table 4: Top 20 sites title of total tags: sites with many total tags and little new tags

investigate and verify the user's awareness.

5.3 SITES WITH LITTLE AMOUNT OF TOTAL TAG, AND MANY NEW TAGS

Finally, we will verify (3)the sites with little amount of total tags, and many new tags. On Table.8, that shows the top 20 of the total tags, internet related contents were at number 10, and marketing related at number 19, each being included once.

The top vocabulary that was used in these sites (Table.9) were mainly dealing with hobbies and culture, such as "movie"," entertainment", "music", "world cup (football)", "football", and "sports". The subdivisions of these various genres were based on the individual's interest and preferences. It is also a field where it constantly transmits the newest trends. The vocabulary in the 3rd place was "new products". We can consider that the contents in these sites are evaluated as new by individual's interests and preferences.

6 CONCLUSION

6.1 SUMMARY OF ANALYSIS

In this research, by analyzing the accumulated data and metadata in the SBM, we attempted to develop a new information search method. By initiating implemented metadata equal to tagging history time series analysis, and using a new tag ratio as an indicator, we classified the accumulated sites of the SBM into the following 3 categories, not only to extract interest and preference of the user community, but also show us the possibility to extract contents based on our target in accordance with the intended use.

(1) Sites with many total tags, and little amount of new tags
⇒Contents centered on active user's interests and concerns

(2) Sites with many total tags, and many new tags
⇒Topics related to medium of communication between the users

rank	tag	number of usage	rate(%)	accumulation (%)
1	Google	26	3.75	3.75
2	YouTube	21	3.03	6.77
3	SNS	17	2.45	9.22
4	Web2.0	16	2.31	11.53
5	Yahoo!	14	2.02	13.54
6	advertising	12	1.73	15.27
7	video (in Japanese)	11	1.59	16.86
8	CNET	11	1.59	18.44
9	Yahoo	11	1.59	20.03
10	search	10	1.44	21.47
11	Mobile	9	1.30	22.77
12	EC navi	8	1.15	23.92
13	Google(in Japanese)	7	1.01	24.93
14	web	6	0.86	25.79
15	video (literal)	6	0.86	26.66
16	video delivery	6	0.86	27.52
17	ITmedia (Japanese ITC news site)	6	0.86	28.39
18	mobile phone	5	0.72	29.11
19	search engine	5	0.72	29.83
20	ranking	5	0.72	30.55

Table 5: Top 20 tags: sites with many total tags and little new tags

(3) Sites with small amount of total tags, and many new tags
⇒Contents that were evaluated as new by individual's interests and preferences

6.2 POTENTIAL OF WAY TO SEARCH ON PARTICULAR PURPOSE

SBM is mainly used for recognizing latest topic among the users' community by ranking on the top page, or searching useful contents screened by other users now. However, the result of analysis bring the possibility that you can select way to find contents to meet one's end among 1) to catch up with the latest fashion of the community, 2) to find contents being useful for medium of communication, or 3) find novel topic faster. The value created by SBM is extended from just the function of filtering contents based on wisdom of crowd (Surowiecki, 2004) to vitalization of the community by provide topic to stimulate communication among the members.

6.3 REMAINED TASKS

At present, this research shows the results from only analyzing the accumulated data in the SBM. There is a need to analyze results in accordance with the interest and concerns of the users, and to consider on what degree users' loyalty regarding the SBM platform is formed through user's survey. Through these surveys, by analyzing from various aspects the mechanism of gathering loyalty SBM will accumulate data and we believe that this can lead us to the understanding of the sophisticated tagging function and the developing of its application method. Our research was an unprecedented attempt to use SBM's actual data and develop a versatile information search method. We need to analyze the user's information search organization activity and behavioral psychology of communication to understand the effect of the platform design. Also the tag, etc and characteristics of the metadata was not fully analyzed. With these perspectives, and as the understanding for the SBM at all levels advance, making it possible for the method we propose in this research to be brushed up.

Finally, we would like to start by thanking Toshiyasu Tsushima and Akiyo Sugawara from EC Navi Company, for providing the data that was the foundation of our analysis, and everyone from EC navi labs: Buzzurl development operation representative. With this data, we were able to apply this actual data of SBM, which there are only few cases world wide, to develop a method.

rank	title of site	number of all tags	new tag ratio(%)
1	Cyber Agent inc. hire Horie (ex-CEO of Livedoor inc.) as CTO: nikkansports.com	121	52.9
2	ranking of popular flavor of "Umai-bo"(Japanese molded confection) - goo ranking	77	68.8
3	list of software which adverse Windows OS works - GIGAZINE	72	61.1
4	Whit is your favorite movie directed by Hayao Miyazaki?	71	57.7
5	the last episode of Doraemon in catoon	67	68.7
6	CEO of EC navi compare Kakaku.com to Yahoo!, EC navi to Google - CNET Japan	66	51.5
7	EC navi redesign SNS service - CNET Japan	65	50.8
8	Opt developed new way of creating advertising on the basis of classificating consumer in 3 segments - CNET Japan	59	55.9
9	"SaaS" "social DB" "long tail": keyword of Drecom's strategy of development - CNET Japan	58	55.2
10	jig.jp provide tool to modify schedule in mobile phone via PC	55	67.3
11	ITmedia News : Cyboze Inc. to IPO in the first section of the Tokyo Stock Exchange	54	72.2
12	Amazon Japan start commission sale - CNET Japan	53	58.5
13	nakata.net – official site of Hidetoshi Nakata	52	57.7
14	Netage group's use new brand name "Saaf" for all service line up and integrate user account with RSS Ads, Inc - CNET Japan	52	59.6
15	Japan.internet.com Web business - how to title and write explanatory text	52	75.0
16	EC navi / Vote; Do you use mixi ?	51	54.9
17	ITmedia News : NTT break into video share service like YouTube	51	51.0
18	careers to study appropriate fee for connecting the Net - CNET Japan	49	77.6
19	Hatena inc. to foray into U.S and develop services in English: IT-PLUS	47	70.2
20	scene like fiction	47	66.0

Table 6: Top 20 sites of total tags: sites with many total tags and many new tags

References

Davenport, T. H. and Beck, J. C. (2001). *The Attention Economy: Understanding the New Currency of Business*. Harvard Business School Press.

Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., and Tomkins, A. (2006). Visualizing tags over time. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 193–202, New York, NY, USA. ACM Press.

Jones, W., Phuwanartnurak, A. J., Gill, R., and Bruce, H. (2005). Don't take my folders away!: organizing personal information to get ghings done. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1505–1508, New York, NY, USA. ACM Press.

Marlow, C., Naaman, M., Boyd, D., and Davis, M. (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. *Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 31–40.

Mathes, A. (2004). Folksonomies-cooperative classification and communication through shared metadata. *Technical Report, University of Illinois Urbana-Champaign*.

Ministry of Internal Affairs and Communications, Japan (2006). White book on information and communication technology 2006.

Niwa, S., Doi, T., and Honiden, S. (2006). Web page recommender system based on folksonomy mining. *Journal of Information Processing Society of Japan*, 47.

Ohmukai, I., Matsuo, Y., Matsumura, N., and Takeda, H. (2006). Realizing Community Web Platform. In *The20th Annual Conference of the Japanese Society for Artificial Intelligence*.

Simon, H. A. (1996). *The Sciences of the Artificial*. MIT Press, Cambridge, Massachusetts, third edition.

rank	tag	number of usage	rate(%)	accumulation (%)
1	EC Navi(in Japanese)	11	1.33	1.33
2	neta (troll)	9	1.09	2.42
3	funny	8	0.97	3.38
4	web2.0	7	0.85	4.23
5	news	7	0.85	5.07
6	ranking	6	0.72	5.80
7	EC	6	0.72	6.52
8	entertainment	6	0.72	7.25
9	ECnavi	5	0.60	7.85
10	PC	5	0.60	8.45
11	gourmet	5	0.60	9.06
12	far-out	4	0.48	9.54
13	social	4	0.48	10.02
14	PC(in Japanese)	4	0.48	10.51
15	ITC	4	0.48	10.99
16	advertising	4	0.48	11.47
17	SNS	4	0.48	11.96
18	Blog	4	0.48	12.44
19	Web service	4	0.48	12.92
20	The Internet / ITC	3	0.36	13.29

Table 7: Top 20 vocabularies used on new tags: sites with many total tags and many new tags

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business,Economies, Societies and Nations.* Doubleday.

Tsuda, D. (2004). *Doing Google - Searching Information on the Web by Google; the Search Engine.* Mainichi Communications Inc.

rank	title of site	number of all tags	new tag ratio(%)
1	extinct and threatened specie: Humboldt penguin increase too much in Japan : science : YOMIURI ONLINE	7	71.4
2	BADSTRING	7	57.1
3	【SPA! (magazine for young men in Japan】kuma-eri (a woman arrested on arson)	7	57.1
4	recruiting for government-paid secretary for Taizo Sugimura (House of Representatives member in Japan)	7	71.4
5	kawasaki's dialy - meet shokotan (Japanese idol)	7	57.1
6	for beautiful human life(blog) - spy photo put on the magazine without permission and to abuse human rights of otaku-woman	7	57.1
7	you see something on Travolta's hairline : ABC (Amarican baka commedy) commity	7	71.4
8	remark of arrested announcer sneaking camera shots of the woman	7	85.7
9	Happa-tai get tobe world famous	7	85.7
10	Google to offer free analytics service - CNET Japan	7	57.1
11	Digg View	7	71.4
12	discovered bags in the soft of mobile phone manufactured by sharp inc and to be corrected	7	57.1
13	sites for downlording wallpaper for high difinition monitor - GIGAZINE	7	71.4
14	Emily Chang - eHub Interviews Lingr	7	71.4
15	code.mfac.jp - Trac	7	57.1
16	topics for blog except meal nor readings, you can make lots of entries (Lifehacks)： Nbonline	7	57.1
17	nantara kantara (blog): Pluto,,,,,,,not planet	7	71.4
18	Also the capacity 10G byte appears! –The web mail competition which is converted keenly - CNET Japan	7	71.4
19	access hours for "CGM"; blog, SNS, BBS extends - - Internet advertising workshop investigation	7	100.0
20	Drawing 3D picture with Acrobat Reader	7	100.0

Table 8: Top 20 sites: sites with little total tags and many new tags

rank	tag	number of usage	rate(%)	accumulation (%)
1	movie	47	0.43	0.43
2	google	46	0.42	0.85
3	new product	44	0.40	1.25
4	entertainment	39	0.36	1.60
5	neta (troll)	37	0.34	1.94
6	mobile phone	37	0.34	2.28
7	Microsoft	32	0.29	2.57
8	Blog	32	0.29	2.86
9	social	32	0.29	3.15
10	the Internet	31	0.28	3.44
11	music	31	0.28	3.72
12	game	27	0.25	3.97
13	SNS	27	0.25	4.21
14	Blog	27	0.25	4.46
15	world cup (football)	27	0.25	4.70
16	football	26	0.24	4.94
17	security	25	0.23	5.17
18	sport	24	0.22	5.39
19	computer	24	0.22	5.61
20	ITC	24	0.22	5.83

Table 9: Top 20 vocabularies used on tags: sites with little total tags and many new tags