Classification and Structuring of People and Learning Content by Characteristics of Knowledge and Methods of Expression

- An application to support self-learning and mutual learning -

Shogo KITAGAWA*, Isao SHINOHARA**, Toshikazu KATO*

** Department of Industrial and Systems Engineering, Chuo University 1-13-27, Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan, {Shogo_Kitagawa, kato}@indsys.chuo-u.ac.jp*

*** Kyodo Printing Co., Ltd. 4-14-12 Koishikawa, Bunkyo-ku, Tokyo 112-8501, Japan, i_shinohara@kyodoprinting.co.jp*

Abstract: Services that allow learners to share or use Web content that is helpful for studying (e.g., Hatena Bookmark, del.icio.us) have become available in recent years. In these services, however, learners search for information using data on interest, popularity and relevance that have been extracted from evaluations and other information left by other learners. This makes it difficult for them to find the desired information. In light of this, the authors summarized the problems of current services: ⅰ) learners don't know what content will meet their current needs, ⅱ) learners don't know what content is easy for them to understand, and iii) learners do not know which data they should obtain because they don't have an overall idea of what they want to know. To solve these problems, the authors propose to deal with user and content information from the perspective of "knowledge," which includes information that is contained in the content, knowledge that is needed to understand the content and knowledge that the user possesses. This paper discusses methods for categorizing and structuring learners and learning content by focusing on the characteristics of knowledge and methods of expression.

Keywords**:** *expression of knowledge through media, folksonomy, support for self-learning and mutual learning, personalized service, face-to-face cooperation and support for mutual help*

1. Introduction

The authors launched the "Helping-Hand Mobile" service (previously called "IT Ranger"), in which campus volunteers support mutual learning, for research and development to support face-to-face cooperation and mutual help using mobile phones [1-2]. This paper discusses how information can be shared in mutual learning.

As indicated by the recent emergence of "consumer generated media" (CGM)[3] as a common term, it is now possible for many people to deliver information on the Internet with ease. Also, based on the Japanese Government's "e-Japan" strategy implemented in January 2001, information-oriented school education has been promoted nationwide [4]. These developments have prompted the improvement of Web content useful for studying. Currently, however, users of information-sharing support services tend to find it troublesome to obtain the needed information on the Web, because they have to search using keywords or various other means and browse through much Web content before finding the appropriate content. This is due to the fact that the current systems use data on interest, popularity and relevance that have been extracted from evaluations, tags, links, trackbacks and bookmarks left by other users.

As a solution, the authors propose to handle user and content information from the perspective of "knowledge," such as information that is contained in the content, knowledge necessary to understand the content and knowledge the person possesses. This will work to establish an information-sharing support service that allows users to easily obtain the desired information.

2. Problems with information-sharing support services

The problems with information-sharing support services that make it difficult for learners to obtain necessary information

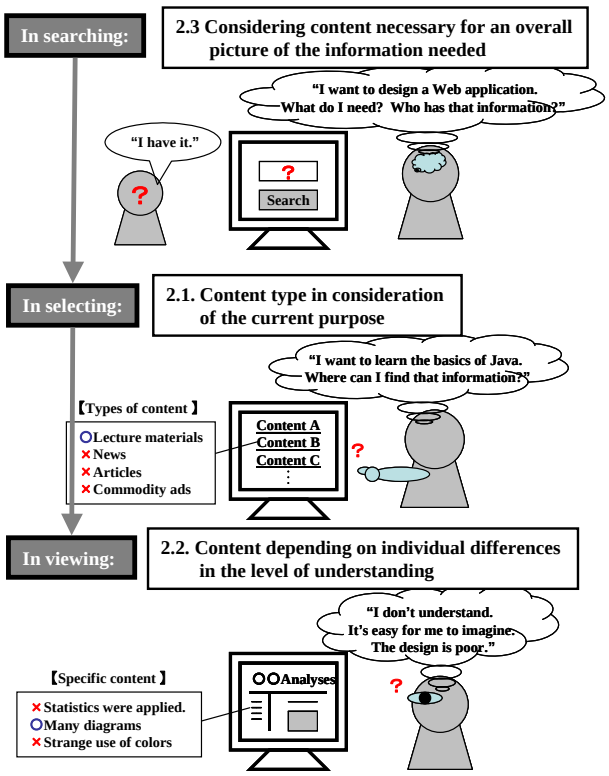

Fig. 1 Problems with information-sharing services for learning

are summarized below in three phases through which learners obtain information (Fig. 1):

- At the start of the search, learners do not know what they need to know to obtain target data (discussed in 2.3 below)
- At the phase of data selection, learners do not know from what content they can obtain the necessary data (2.1)
- At the phase of viewing content, learners find it difficult to understand what they want to know (2.2).

To solve these problems, it may be necessary for providers of the current services to consider the three factors below.

2.1. Considering content in light of the current purpose

When people begin to learn about Java and look for study materials intended for beginners, they tend to obtain additional content that they do not need then, such as related news, ads and articles. To solve this problem, it is necessary for the current services to have a system that considers for what purposes the users are searching the Net, so that users can know specific types of content that contain the necessary information.

2.2. Considering content in light of individual differences in the level of understanding

People who have just begun to learn statistics will have difficulty understanding lectures on applied statistics, for example, while people with mathematical knowledge will find it easier to understand explanations given in the form of equations than in the form of verbal explanations. Thus it is necessary for the current services to have a system that considers personal differences in comprehension and offers users search results that correspond to the level of comprehension.

2.3. Considering content necessary for an overall picture of the information needed

A learner who is beginning to study how to design a Web application, for example, sometimes does not know what he or she should know first, or the name of a desired learning method. This makes it impossible to search for that information. Therefore, it is necessary for the current services to have a syst em that allows users to determine the data that are necessary for searching appropriate Web content in consideration of an overall picture of their needs.

3. Classification and structuring of learners and learning content according to the characteristics of knowledge and methods of expression

In consideration of the three points summarized in Chapter 2, the authors propose a method for classifying and structuring learners and learning content by focusing on characteristics of knowledge and methods of expression based on user evaluation of learning content. The service systems used here are social bookmark services (e.g., Hatena Bookmark, del.icio.us) [5-6]. As illustrated in Fig. 2, this method has the three characteristics

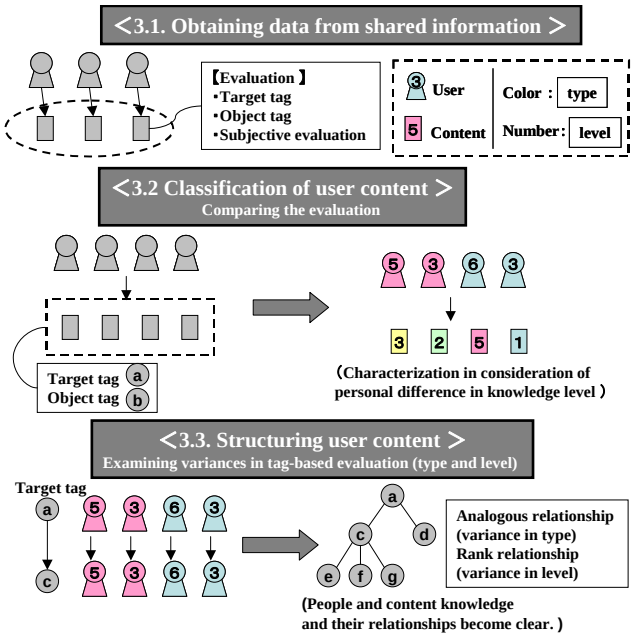

Fig. 2 Structuring of learners and learning content

given from 3.1 to 3.3.

3.1. Obtaining classification tags with rules and subjective appraisals

Users were asked to evaluate Web content by offering two tags with rules and conducting three subjective appraisals. Tagging based on rules offers information necessary for the classification of Web content.

In the current information-sharing services, users can tag any information freely. This provides little information on the content that users have found difficult to understand. In our proposed system, content is classified based on reasons that are indicated by tagging with rules, to allow users to efficiently obtain data necessary to find the target Web content. By searching websites using this system, users can find content that is appropriate for their purposes. These data are also used for the classification and structuring of people and Web content based on individual differences in knowledge level.

Two tags with rules used here are as below. Users can add other tags as “free tags” at their discretion.

・Target tag

This tag shows the target of the Web content (e.g., Java, statistics, principal component analyses) and the specific target of user evaluation within the Web content.

・Object tag

This tag shows types of Web content (e.g., summaries, news, articles, blogs) and enables users to obtain Web content suited to their purposes at the moment of searching. This also indirectly shows the user’s purpose of using the Web pages.

The details of three subjective assessments by users are described here. For more specific analyses, it is also possible to prepare rule-based tags corresponding to these appraisals and make them available for users when necessary.

・Scores indicating level of understanding

This shows how much knowledge the user has and how well the user has understood the Web page (scored from “too easy” to “easy enough” and “too difficult”)

・Scores indicating preference

This shows the level of compatibility between the Web page and the user’s preference (scored either as “preferable” or “not preferable”).

・Scores indicating whether the content is recommendable

This indicates whether the user would recommend the Web page to other users. (Many existing services have already introduced this appraisal. Web pages are scored either “recommendable” or “not recommendable”)

3.2. Classification of people and learning content

The current services simply aggregate evaluation results without considering individual characteristics of users who have evaluated the Web content. This tends to result in a situation in which users find it difficult to read a book that has been recommended as easy.

Our proposed system classifies users by comparing their subjective appraisals of Web pages. Because the specific knowledge possessed by users (“knowledge type”) is reflected in their evaluation tendencies on Web content, it is possible to determine how deeply users understand the relevant information (“knowledge level”) by comparing appraisals between users of similar knowledge types. Therefore, this system classifies people based on similarities in knowledge level and type, while it categorizes Web content in consideration of users’ individual characteristics, namely, by examining their knowledge type and level and their evaluation of Web content. In this way, our system enables users to obtain content suitable for their individual level of understanding while browsing the Web. This system also allows them to acquire content that matches their preferences or that uses expressions easy for them to understand.

3.3. Structuring people and learning content

Since methods for representing people’s knowledge structures ha ve not been established, our system accumulates comparative data on each user’s assessments of Web pages for each tag, and structures learning content based on variances in overall similarities in users’ subjective evaluations. When many of the users, who have been determined to be of the same knowledge type from their evaluation on a certain tag, again show the same tendencies in their evaluation on another tag, then these users are considered to be capable of understanding the tagged content when they have similar knowledge level. This is called an “analogous relationship” here. Meanwhile, when differences are observed in these evaluations, this system determines that the relevant user will not be able to understand this information because he or she does not have that information. These differences express what is called a “rank relationship” here. In our system, people and content are structured based on folksonomy data in consideration of these analogous and rank relationships.

By using existing tag data and in consideration of users’ analogous and rank relationships, it is possible to estimate user evaluations for Web pages that have been tagged but not yet evaluated.

4. Overview of the system and service

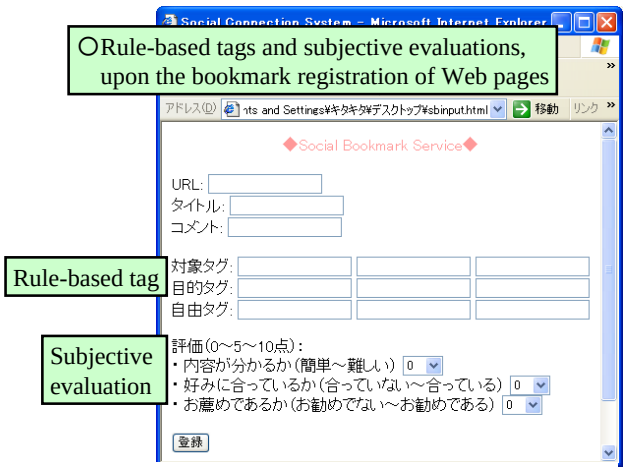

Fig.3 Window for contents registration

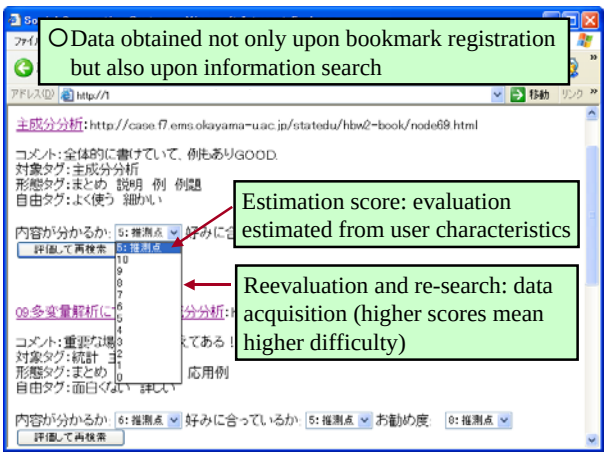

Fig. 4 Window for content search results and re-search

Figures 3 and 4 show the service's windows for content registration and search results (and re-search), respectively. Many of the current services have come to employ user evaluations through rule-based tagging and subjective assessments. Our system obtains data also from user reevaluation of the search results to be used for re-search, to ensure that users can constantly obtain various data and use the service under a system load not very different from that of the social bookmark service. Our system also makes it possible to constantly manage individual user characteristics that are subject to change.

Described below is the process of accumulating user evaluations of Web content and of categorizing and structuring people and content such as to enable users to obtain the data they need.

4.1. Data acquisition

In this system, a set of data consisting of two tags with rules (target and object tags) and three subjective appraisals (knowledge, preference and recommendation) made by users while viewing content is obtained from each user.

4.2. Data analyses

For each tag, users are compared based on scores for understanding and preference. The relevant user (User A) is compared with each other user (User X) through the following process:

- Evaluation data on Web content made by both Users A and X are obtained.
- The difference in evaluation scores given by Users A and X is determined. The differences are averaged, and the mean value is subtracted from all the difference data. The absolute of the value after subtraction is summed up, and this sum is used as an index for similarity in knowledge type between Users A and X.
- The difference in the mean value of evaluations for relevant Web content between Users A and X is regarded as the difference in knowledge level.

(The structuring of people and learning content is to be discussed after the validity of the proposed classification has been verified.)

4.3. Content selection

The system uses the following methods to choose content based on similarity in knowledge type and differences in knowledge level, by using data obtained for the comparison of users.

- Selecting users who are considered close in type based on scores for understanding and preference (specifically, users of similarity level 5 or below are considered close in knowledge type. If there is no one applicable, the top 10 users are used).
- By focusing on differences in scores for understanding and preference in the evaluation of Web content by the selected users, Web content that the relevant user (User A) can understand moderately well (score of about 5) and content that agrees with his/her preference (score of over 5) are obtained.
- The scores of recommendation for the Web content are weighted as follows. The weighted score is regarded as an index for understandability, and information with higher index is offered to users.
  - Users can choose Web content with a higher score for preference or for understanding, depending on the purpose of use.
  - Evaluation of users with knowledge level higher than User A is given greater weight. In this system, recommendation of users with more knowledge than the relevant user is considered to be a more useful reference.
  - Evaluation of users of lower preference scores than User A is given higher weight. In this

system, recommendation of people with stricter preferences than the relevant user is considered to serve as a useful reference.

5. Application to face-to-face cooperation

Information obtained by classifying and structuring people and content from the perspective of knowledge can be applied not only to information-sharing but also to support for face-to-base cooperation. Figure 5 shows problems and concepts of such support services.

Content to which the service can be applied includes, for example, that related to visualization of knowledge, assistance in building learning communities, matching between supporters and clients in accordance with knowledge level, and support for instruction in consideration of knowledge level.

We intend to offer a support service for the mutual help of users by linking them and enlarging their linkages through this service. This service is also expected to develop as a support service for mutual help to ensure that anyone can contribute as a supporter, depending on ability, and can grow together by mutual instruction.

6. Verification of the validity of the proposed support system for mutual learning

Two preliminary experiments were conducted on the structure of the proposed mutual study support system.

6.1. Necessity of the proposed method

To verify the necessity of the proposed method for structuring learners and learning content, 11 laboratory members served as subjects and subjectively evaluated Web content (scoring of understanding, preference and recommendation). The tendencies in their evaluations were examined. The Web

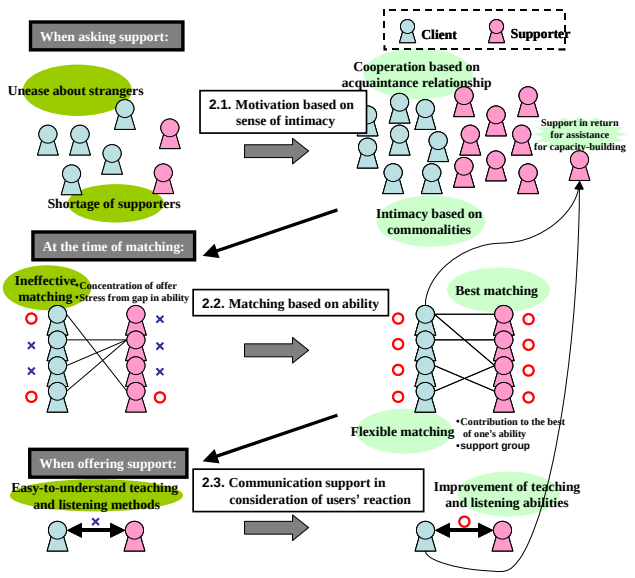

Fig. 5 Problems and concepts of mutual help support service

Table 1. Content evaluation result

		Subject 1	Subject 4
Web page B	Score for understanding	2	9
	Score for preference	1	7
	Score for recommendation	1	10
Web page E	Score for understanding	7	0
	Score for preference	7	3
	Score for recommendation	8	1

pages used for this evaluation consisted of those the authors had determined were written on one of four different subjects (Internet scams, Internet security, multiple linear regression analysis and principal component analysis). Five pages were used for each of the four subjects, forming a total of 20 pages.

Excerpted results are shown in Table 1. According to this table, significant differences in evaluation are observed between Subjects 1 and 4 on two Web pages about principal component analysis. As for Web page B, the score given by Subject 1 was significantly low, whereas that given by Subject 4 was considerably high. The reverse was true for evaluation points on Web page E.

Because individual differences in evaluation also tend to be demonstrated even among laboratory members with similar characteristics, it is considered necessary to classify and structure people and content in light of personal differences in knowledge and preference when preparing information-sharing services.

6.2. Needs for learning through mutual help

To clarify the needs for mutual learning support services, the authors conducted a questionnaire survey in June 2006 on 142 first-year students who were taking an information technology course at our university. To each question, they assigned a score from 1 to 5 (1: yes, very much; 2: yes; 3: difficult to say yes or no; 4: no; 5: not at all). The questions and results are as shown in Tables 2 and 3. The results indicate the following:

(1) To questions 3 to 11, almost all subjects gave score of 1 to 3, and the average score was below 2.5. From these results, it is considered that most people who demand various things as clients have the potential to become supporters. Therefore smooth cooperation among users will be possible by estimating users' knowledge levels and promoting efficient matching among users based on this method, as well as by offering support.

(2) Almost all subjects gave scores of 1 to 3 to questions 3 to 6. This indicates that teaching assistants (TAs) are expected to offer face-to-face or cell-phone support to students before or after classes on content other than lesson content.

(3) Questions 7 and 9 were given similar scores, while question 8 received more 1s and 2s than questions 7 and 9. These

Table 2: Ouestionnaire items

Table 2. Questionnaire item

1. Do you think you're keeping up with this class?
2. Do you think you've understood the lesson thoroughly. In other words, do you think you're able to answer questions and apply the content to other fields?
3. Would you like graduate students who support your studies during the class (hereinafter: "teaching assistants" (TAs)) to assist you in studying this subject before or after class?
4. Would you like TAs to assist you on matters other than those about this class (problems in studying on their own, campus life, etc.)?
5. Do you think it would be helpful if you could ask for support not only by PCs but also by mobile phone?
6. Do you also hope for support not only by mail or phone but also face to face?

●Please answer the following questions, which aim to examine how willingly you will support people in trouble.

7. Would you offer support to anyone in need, even if there is no return?
8. Would you offer support to someone in trouble who has helped you before?
9. Would you offer support to an acquaintance of someone who helped you when you were in trouble?
10. Do you think you'd be willing to offer support to anyone in need after you were helped by someone?
11. Would you support anyone in trouble if doing so would guarantee that you would also be helped when in need?

Table 3. Questionnaire results

No.	1	2	3	4	5	6	7	8	9	10	11
No. of 1s	5	3	43	31	41	37	28	70	31	41	53
No. of 2s	33	9	56	53	46	59	82	65	83	67	66
No. of 3s	47	40	37	37	36	37	24	6	23	27	18
No. of 4s	49	69	3	17	14	6	7	1	3	5	3
No. of 5s	8	21	3	4	5	3	1	0	2	2	2
Avg.	3.15	3.68	2.06	2.37	2.27	2.15	2.09	1.56	2.03	2.01	1.84
Variance	0.93	0.77	0.84	1.08	1.19	0.86	0.62	0.38	0.59	0.75	0.69

indicate that users tend to prefer helping acquaintances (users whom they have met before or know personally) rather than helping total strangers. Therefore, a system to encourage mutual support based on acquaintance relationship may be effective.

(4) The mean evaluation scores for questions 7, 10 and 11 are 2.1 or below, which may indicate that users tend to think what they have done for others will be eventually paid back to them. They are also willing to support people who have helped them and are optimistic about assisting others when it is assured that they will be helped when in need. Therefore, it may be effective to incorporate in this service a system that encourages chain reactions of mutual help among users by introducing appreciation points.

7. Conclusion and future prospects

To examine information-sharing in mutual learning, this paper has summarized problems and matters to be considered in three points and has proposed a method for classifying and structuring learners and learning content by focusing on characteristics of knowledge and methods of expression. With the preliminary experiments, the need for face-to-face cooperation support service was verified.

This system could be applied to business and nursing services, and it could be used in the evaluation of cooperative activities and as criteria for judgment of one's ability to offer support.

As a next step, it is necessary to systemize and implement this method toward verifying and improving its appropriateness.

Acknowledgements

This research is partially supported by Grant-in-Aid for Scientific Research (S), No. 19100004, Japan Society for thePromotion of Science.

This research is partially supported by Grant-in-Aid for Scientific Research (S), No. 19100004, "Robotics modeling of diversity of multiple KANSEI and situation understanding in real space", Japan Society for the Promotion of Science.

References

[1] Emi Sueyoshi, Isamu Shinohara, Shunichi Kato. Support Service for Local Communication Using Mobile Communication. Japan Society of Kansei Engineering. Tokyo, 2004.

[2] Satoshi Tanaka, Kaori Komiya, Isamu Shinohara, Shunichi Kato. Construction of Satisfaction Presumption Model Based on Social Connection and Information Search Algorithm – Rescue Service Based on Social Connection. 7th Meeting of Japan Society of Kansei Engineering, 2005; 181.

[3] "What is Web 2.0?" Tim O'Reilly http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

[4] White Paper of the Ministry of Education, Culture, Sports, Science and Technology, Part II Chapter 11, Topic 1 "For Further Promotion of IT Application to Education," http://www.mext.go.jp/b_menu/hakusho/html/hpba200501/002/011/topics01.htm

[5] Hatena Bookmark, http://b.hatena.ne.jp/

[6]Del.icio.us, http://del.icio.us/popular/