Research of Information Organization based on Folksonomy

B.G.YU & J.W.GU

School of Management; Key Laboratory of Process Optimization and Intelligent Decision Making of Ministry of Education, Hefei University of Technology, Hefei, China

ABSTRACT: This paper proposes a method which uses RDF to express resources, tags and users of Folksonomy and allows users to assign tags freely by achieving the storage and query of RDF data. It provides a means for completely open and sharing of information between users. And it is also applied in the manufacture enterprises' website that it widens the application range of folksonomy. The method can not only improve the degree of users' participation, but also realize the efficient management of mass data and the quick query of Information.

KEYWORD: Folksonomy; RDF; Information Organization; Information Representation

1 INTRODUCTION

With the rapid development of science and technology, the way of online trading prevails, going with mass information and many counterfeit products which cause problems in product quality. The enterprise official website, as the access to information of products and the platform for purchasing products by users, has the higher credibility. However, most information representation of products given by manufacturing enterprises' websites is usually two-dimensional or three-dimensional, and the information is not well organized or used and is not convenient for users to gain more useful information rapidly. The generation of web2.0 gradually dilutes the boundary between information providers and users. Users are no longer just recipients of information, but also the main body of information generation, sharing and dissemination. Its prominent position is increasingly highlighted. Therefore, manufacturing enterprises needs the better way of information organization and representation to fully enhance the enthusiasm and participation of users, allowing users to make decisions quickly and easily. The user-oriented folksonomy, formed by bringing tags together and organizing public tags, provides a new perspective to solve this problem.

2 RESEARCH STATUS

Folksonomy is generated and used by users, with the purpose of indexing and retrieving dynamic network resources more conveniently. Research abroad began in 2005, and has reached a certain height both in theory and applications. The main research focus lies in nine areas: information management, metadata, semantic Web, collaborative tagging, knowledge management, knowledge sharing, search engines, tag mark and sorting. Ali Shiri [1] compared and analyzed interfaces, features and functions of 10 social tagging systems, and discussed the way of user interface encouraging and providing label distribution, label exploring, label browsing and using the label when interacting with social websites. Ikki Ohmukai [2] et al proposed a social tagging system, which uses multiple metadata and personal network to build a community-based Ontology, so that users can map their own tags to different tags from different friends to associate with other users. Folksonomy was early applied to libraries and sites abroad, and has got a great development. But China started relatively late, and most research is according to foreign research. For examples, document [3] introduces and discusses basic content of Folksonomy; document [4] and document[5] respectively discuss the contrast and the integration between Folksonomy and Ontology [6]; document [7] studies the semantic association enriching of Folksonomy [8]; documents [9-11] introduce and analyze some sites about Folksonomy such as Flickr, Del.icio.us, Douban.com, CiteULike, etc. And they make a compare among these sites. In addition to some big sites have applications about Folksonomy (such as Dangdang, Douban.com, etc.), most sites have not applied yet. [12]

3 OVERVIEWS AND CONCEPTUAL MODELS OF FOLKSONOMY AND RDF

3.1 *Folksonomy*

The word 'Folksonomy' was first proposed by the United States Information Architecture expert Thomas Vander Wal in August 2004, fusing folk and taxonomy. Thomas defines Folksonomy: It's the result of tagging any information and object owned URL freely by individual for facilitating the retrieval. The tag is formed in the social environment, and it opens to others and shares with others. The behavior of tagging is the process of digesting and absorbing information for users. [4]

In this paper, quoting the thought of Folksonomy is for its openness, sharing and the reinforcement of user's participation. Folksonomy mainly consists of three basic elements: resources, tags and users. For now, the most use and complete of its application is the social tagging system.
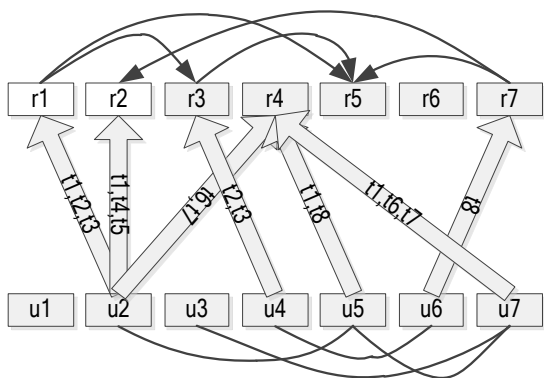

Figure1. the Model of Social Tagging System

Figure1 shows a conceptual model of the social tagging system. In this model, users can tag the particular resources, but the users' information space is neither closed nor isolated, but rather open to sharing. Users can know evaluation of a resource from other users by viewing other tags tagged by them. Users can also find community members with common interests through tags. In Figure1, connecting lines between users express relationships between them. Users can add any tag to any resource. That is to say, one resource can be tagged by different tags and one tag can be used to tag different resources, so that different resources can be associated and that is represented by the connecting lines between the resources in Figure1. Resources, tags and users can be connected through the behavior of tagging.[13]

3.2 *RDF*

After browsing extensive documents, it is found that resources, tags and users in Folksonomy can be represent in RDF. And the specific application of

Folksonomy can be achieved by the way of achieving storage and query of RDF.

RDF, which is called Resource Description Framework, is a common language proposed by W3C for describing Web information. It is a new standard after XML and is the XML document defined metadata. Any object owned a URL as a unified identity in RDF is resource, which can not only refer to the object indicated and contained a URL in the web, but also refer to the property of the object. The statement of resources' properties is description which shows the characteristics of resources, points out the linkages between resources and greatly increases scalability. Framework is a generic model and is nothing to do with the resources described.

As a kind of information representation language, RDF proposes a simple model to represent data: triple (S, P, O), which is used to describe Web information, where S means Subject, P means Predicate, O means Object. It is shown in Figure2 (b). From the point of view of the diagram, data can also be represented by a directed graph model with marks, named RDF graph. The source node means Subject, the destination node means Object, and the directed line segment connecting the subject node and the object node means Predicate, as shown in Figure2 (c) below. Thus, it is convenient to describe resources and relationships between them.[14-15]
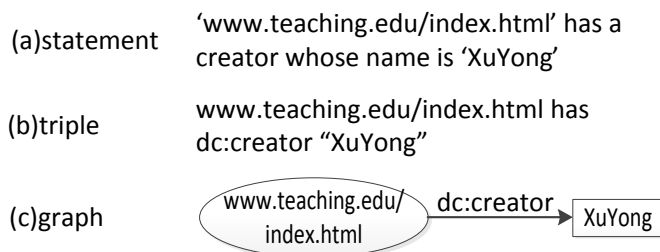

Figure2. the correspondence among the statement, triple and graph model of rdf

4 INFORMATION ORGANIZATION AND REPRESENTATION BASED ON FOLKSONOMY AND RDF

4.1 *Information representation based on Folksonomy and RDF*

The product information representation given by manufacturing enterprises has intricate relationships with the customer, and it is the key reference for their decisions. But for the moment, most manufacturing enterprises lack a high value on their product information representation. Like HEFEI MEILING CO, LTD, whose product information is shown in Figure3.
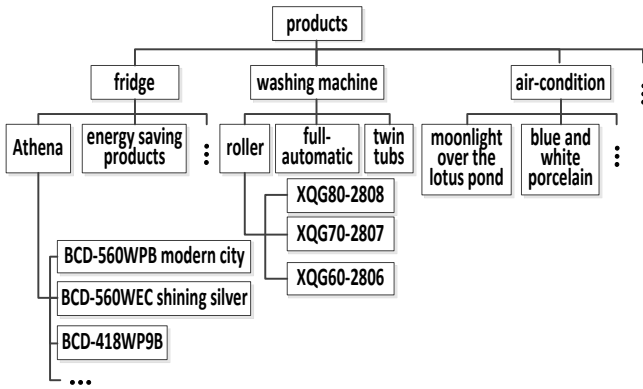
Figure3. the Product Information Representation of Meiling

The company achieves its three-dimensional representation by three properties which are product types, series and models. Display specific information after turning to the specific product. Products' specific information is divided into five parts: the purchase information (including models, inventory, price, category, add product tags, etc.), product attributes (including series, color, display format, weight, etc.), description of goods, purchase records and users' evaluation (including users' names, emails addresses, ranks and comment content). Such kind of information organization and representation is of relative limitations. The query provided is based on the precondition that users know product models. Therefore it is not conducive for users to know product information user-friendly and fast to choose product met their own needs.

As a major platform for current online shopping, Taobao supports to select some properties to achieve further filter the products on the basis of similar products in the three-dimensional representation. In addition, it adds function of keyword matching queries by the use of product names (or so-called product overviews) achieved. But it usually makes merchants extend products' names as much as they can to show their products in matching result. And it leads to that customers often get a large number of irrelevant product information or the information not so in line with their requirements. Moreover, all of the information is provided by the merchant, and customers are only recipients without participatory.

For the convenience of users to make better use of information to enhance the users' initiative and to facilitate the users to make decisions, this paper introduces the idea of Folksonomy, while increasing users' participation, to achieve multi-dimensional representation of product information, so that it can make information be better organized and managed, as Figure4.

Relations among resources, tags and users can be expressed in RDF. It is not only simple and clear, but also helpful to further organization and management of information to achieve the purpose of the realization of Folksonomy applications. Figure5 and Figure6 are part of RDF representation

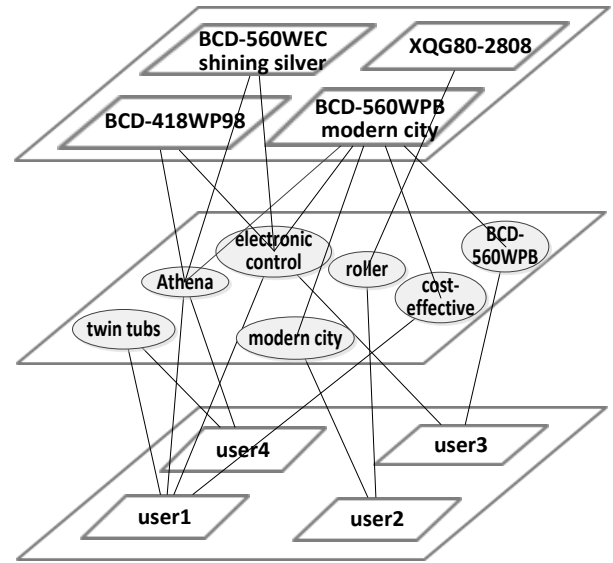of product "BCD-560WPB" modern city tagged by users and its RDF graph representation.

Figure4. the Modified Product Information Representation of Meiling

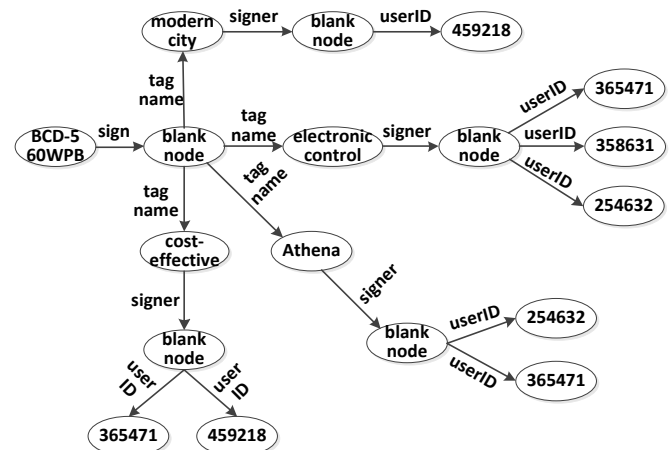Figure5. RDF representation of resources, tags and users

Figure6. RDF graph representation of resources, tags and users

In addition, enterprises can also give basic knowledge related to products (such as classification of refrigerators, types and characteristics of a single door direct cooling systems, etc.), so that users can search by tags, and get a further understanding of products.

4.2 *Information storage based on Folksonomy and RDF*

When storing RDF, each "resource - tag - user" can be viewed as a record. So a record may contains multiple predicates and objects. If using a relational database to store, there will be a lot of duplicate content between different records which is not conducive to efficient storage and query. Given that there are variable number of predicates and objects under the same subject, this paper chooses database Cassandra. It is the expansion of the concept of Bigtable which is sparse and distributed non-relational database. On the basis of row key, column key and timestamp, it defines the concept of column family, used to separate logically and link to the same kinds of data. Its most basic data structure is a column composed of one name/value pairs. A column family contains a number of name/value pairs, and a super column family contains one group or several groups of columns with particular relationships. Figure7 is the data model of Cassandra.
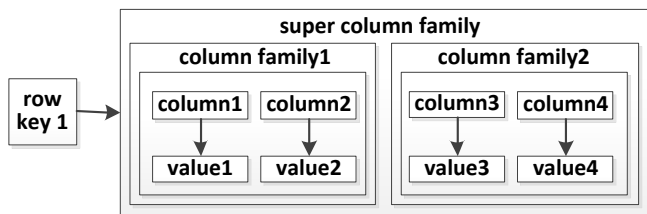

Figure7. the Data Model of Cassandra.

When carrying out the storage, stores subject S with row key, predicate P with column family and object O with column key. And store three times in accordance with 'SPO', 'POS', 'OSP' sequence so that different table stored correspond to related indexes. For different requests, query in the corresponding index structure and access to relatively high query performance by the way of spatial exchange. It is able to quickly locate the corresponding content if giving any element or any combination of each of resources, tags and users. And that is with the help of Links among S, P and O, and with the use of nested key mapping. Table 1 is the corresponding relations among resources, tags and users formed by correlating S, P and O in the data model of Cassandra.

Table1. the Corresponding Relations of a Resource and its Tags and Users

resources	tags	users
http://mall.meiling.com/goods.php?id=364	electronic control	365471
		358631
		254632
	cost-effective	365471
		459218
	Athena	254632
		365471
	modern city	459218

4.3 *Information Query based on Folksonomy and RDF*

Due to the insufficient function of the Cassandra query language CQL, it may not meet the efficient query of 'resources, tags and users'. SPARQL[16], as the search query language for RDF data released by W3C is a good choice. It achieves efficient query functions by the Graph Pattern match. The simplest mode is triplet, similar with triplet of PDF, except that the triplet mode allows the subject, predicate or object to be a query variable, and eventually forming a basic graphics mode by merging triples mode. Blank nodes in the query are viewed as a form of variables. Matching triplet mode with graph pattern is not complicated. Bind variables queried and RDF vocabulary, replacing the variable into the corresponding RDF vocabulary, a successful match graph triple is got. According to different storage and indexing schemes, the variables appear in different positions can choose to use different indexes to carry out the nested key mapping query. Figure8 shows the triplet pattern query and their corresponding indexes available.
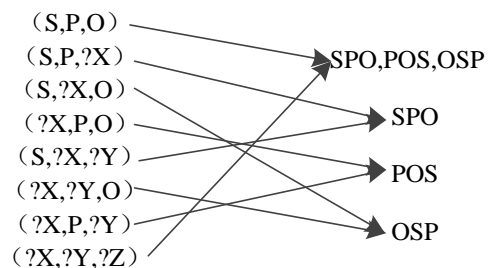

Figure 8. the Triplet Pattern Query and Their Corresponding Indexes Available [17]

Figure 9 and Table 2 is an example of the query and its result above, finding product models and their corresponding users tagged as 'electronic control'.

```
PREFIX rdf:< http://www.w3.org/1999/02/22-rdf-syn-lax-ns#>
PREFIX dc:< http://purl.org/dc/elements/1.1/>
PREFIX t:< http://description.org/schema/tag#>
SELECT ?rdfs:model,?dc:userID
WHERE {?Production t:tag _:tag.
       _:tag dc:name electronic control.}
```

Figure 9. Query of SPARQL

Table2. Results of Query

model	userID
BCD-560WPB	365471
BCD-560WPB	358631
BCD-560WPB	254632
BCD-560WEC	658216
BCD-418WP9B	572689

5 CONCLUSION

This paper expresses resources, tags and users in Folksonomy in RDF, making the thought of tagging by users with freedom in Folksonomy and massive data storage of RDF integrated. And it is simple and clearly shows the relationships among resources, tags and users. It achieves the application of Folksonomy by implementing RDF data storage and query which provides a means of completed information openness and sharing. And it is applied to the manufacture enterprises' websites, and it broadens the scope of application of Folksonomy. Meanwhile, the use of storage thrice solved the problem that it can only start query by row key in Cassandra that can't meet the lack of querying resources with tags. In the full increase users' participations, while achieving the efficient management and fast query of information of large-scale data.

REFERENCES

[1] Ali Shiri. An Examination of Social Tagging Interface Features and Functionalities: An Snalytical Comparison. Online Information Review. 2009, 33(5):901-919.

[2] Ikki Ohmukai, Masahiro Hamasaki, Hideaki Takeda. A Proposal of Community-based Folksonomy with RDF Metadata. http://itu.dk/people/martynas/Thesis/Articles/A%20Proposal%20of%20Community-based%20Folksonomy.pdf

[3] Yu Jinxiang. Folksonomy and Foreign Research Progress. Library and Information Service. 2007, 51(7):38-40.

[4] Wang Cuiying. A Comparative Study of Ontology and Folksonomy. Library Development. 2008, (5):85-88.

[5] Zhang Zhongyun. Semantic Relations Presented in Folksonomy: Folksonomy Local Ontology. Library and Information Service. 2012, 56(20):127-131.

[6] Hend S.AL-Khalifa, Hugh C.Davis. Folks Annotation: A Semantic Metadata Tool for Annotating Learning Resources Using Folksonomies and Domain Ontologies. Innovations in Information Technology, 2006. IEEE, 2006:1-5.

[7] Wei Lai. Review of Rich Semantic Study of Folksonomy abroad. Information and Documentation Services. 2010, (3):40-44.

[8] Lu Jianjiang, Zhang Yafei, Miao Zhuang, Zhou Bo. Semantic Web Principles and Techniques. Beijing: Science Press, 2007:30-65.

[9] Shi Hao, Li HongJuan, Lai Wen, Zhao Ying. Users classification study based on Folksonomy tags. Library and Information Service. 2011, 55(2):117-120.

[10] Xu Jing, Lu Zhangping. Information Organization and Optimization Based on Folksonomy. New Century Library. 2011, (4):34-37.

[11] Li Aiguo. Comparative Analysis of Three Academic Tag System. Library and Information Service. 2007, 51(2):114-116.

[12] Guan Fengting. Analysis of Folksonomy Research Status Based on Bibliometrics. Library Tribune. 2012, 32(4):94-100.

[13] Cameron Marlow, Mor Naaman, Danah Boyd, Marc Davis. HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, ToRead. Proceedings of the seventeenth conference on Hypertext and hypermedia. ACM, 2006:31-40.

[14] Pascal Hitzler, Markus Krotszsch, Sebastian Rudolph. Technology Base of Semantic Web. Yu Yong, etc. Bejing: Tsinghua University Press, 2012:12-73.

[15] Shi Xuelin, Zhao Ying. Semantic Grid Data Modeling and Retrieval Based on RDF. Computer Applications. 2009, 28(9):2324-2327.

[16] SPARQL Query Language for RDF. W3C Recommendation 15 January 2008. http://www.w3.org/TR/rdf-sparql-query/#termConstraint.

[17] Wang Xin, Feng Zhiyong, Du Pufeng, Rao Guozheng, Jiang Yang, Li Xu, Sheng Zhaohua, Jiang Longxiang. Jingwei: A Distributed Large-scale RDF Data Server. Computer Research and Development. 2011, 48(Suppl): 451-455.

[18] Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation 10 February 2004. http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/#section-Datatypes-intro

[19] Jin Hai, Yuan Pingpeng. Semantic Web Data Management Technologies and Applications. Beijing:Science Press, 2010: 162-193.

[20] Teng Guangqing. Knowledge Organization of Digital Library Based on Concept Lattice. Jilin: Jilin University, 2012.

[21] Jiang Longxiang, Wang Xin, Li Xu, Feng Zhiyong. Distributed Storage Solution for Large-scale Semantic Data of RDF. Computer Applications and Software. 2011, 28(11):30-32.