A Method for Enterprise Knowledge Map Construction Based on Social Classification

Lu Liu, Jing Li, and Chenggong Lv

School of Economics and Management, BeiHang University, 100083
Beijing, China,
liulu@buaa.edu.cn, lijing0622@buaa.edu.cn, lcg@sem.buaa.edu.cn

Abstract. Knowledge map is an important and effective instrument of corporate knowledge management. A great number of unordered knowledge resources in the enterprises bring about difficulties to the knowledge map construction. This paper proposes an effective method based on social classification to organize the enterprise knowledge resources and to construct the knowledge map further. According to the characteristics of knowledge usage in the enterprise, the enterprise knowledge map is defined as a domain-centered system to display knowledge and their relationships. By virtue of the collaborative nature of social classification, the proposed method collects individual knowledge tagging data and selects some important topics from tags to form a domain knowledge map. Then topics in different domains are related to each other by similarity and the enterprise knowledge map is constructed. At last, a prototype knowledge map system, which is being implemented for an enterprise, is presented.

1 Introduction

With the acceleration of global economic and IT development, enterprises today endeavor to explore better approaches to improve organizational adoption, survival and competence in the new business environment characterized by dynamic, discontinuous and rapid pace of change [1]. Knowledge management is increasingly viewed as a crucial factor for organizational sustainable competitive advantages [1, 2]. It embodies organizational processes that seek synergistic combination of data and information processing capacity of IT, and the creative and innovative capacity of human beings.

Empirical studies have shown that while organizations learn and create knowledge, they also forget (i.e., do not remember or lose track of the acquired knowledge) [3, 4]. Many organizations accumulate a large amount of knowledge

along with their business progress. Unfortunately, their employees have to spend so much time and effort on knowledge searching and selection before they can find out what they want. Therefore, it is very important to create knowledge map to specify the captured knowledge and relationships in order to facilitate knowledge navigating and searching.

Currently, the enabling technologies of knowledge map mainly are Intranet-based software solutions, which combine powerful visualization techniques with database management system [5]. Yet, while the technological implementation can lead to a useful knowledge map artifact, the process of mapping is even more challenging. Essentially, the knowledge mapping process is the process of knowledge organization and classification. Many organizations, for the convenience of storing and searching knowledge, have their own approaches and practices on knowledge classification, in which they usually design taxonomies manually to organize the key files. Manually organizing knowledge to categories requires significant time and effort, so it is difficult to design taxonomy to cover the whole domain of interest in enough detail. In recent years, automatic document classification has been widely researched and used [6]. This kind of methods has shown to be reasonably effective when dealing with huge amount of various contents on the Internet. However, it is not suitable for organizing the knowledge in a real enterprise environment. A Knowledge map for an organization has to reflect the business process and be comprehensive to knowledge workers, which can not be resolved by current automatic classification methods.

In this paper, we present an improved social classification-based method for knowledge organization and knowledge map construction. The proposed method develops enterprise knowledge map as a multi-level system including the individual level, team level and organizational level. According to the similarity of knowledge structure in a business domain, our method defines domain as the basic organizing unit of enterprise knowledge map to display knowledge and their relationships. Owing to the collaborative nature of social classification, the proposed method collects individual tagging data and selects the important topics from tags to form a domain knowledge map. Then we can relate topics in different domains by similarity and construct the enterprise knowledge map.

The rest of this paper is organized as follows. Section 2 examines related work and background concepts that we have employed. Section 3 presents the outline of the proposed method for knowledge map construction, including its architecture and basic procedure. Then in section 4, the detailed method is provided. In section 5 a prototype system for an enterprise is presented. Finally a conclusion with future research is given in section 6.

2   Related Work

In this section, we briefly present some research literatures related to knowledge map construction and social classification.

2.1 Knowledge Map Construction

A knowledge map is the display of acquired knowledge and relationships [7, 8]. The knowledge in knowledge map may involve various shared contents, such as text, graphics, videos, models and data. The relationships among them are determined by linking concepts or topics discovered from these shared contents. There are several main strategies for constructing knowledge maps, including building directories, manually drawing concept map/topic map, and automatic knowledge classification [9].

A directory is an alphabetical or classified list of names, addresses and other data. On the Internet a web directory is usually used to organize topics in groups and subgroups, such as YAHOO! and Open Directory Project. It is a simple but effective way to organize a large volume of information, especially when coupled with a search engine. Nevertheless, it is recognized that the interface of directory becomes increasingly difficult for users to navigate as the hierarchy grows larger [10, 11].

Concept Map [12] and Topic Map [13] are drawings, in which blocks represent concepts, topics or things and connecting lines represent relationships. This kind of maps can help better organize, display and understand knowledge. However, the creation process requires highly creator's cognitive skills and significant time and effort.

More recent works construct the knowledge maps for some kinds of web contents, employing machine-learning algorithms to cluster the web documents. Document representation and clustering technique are two major issues in text clustering. The vector space model (VSM) is usually adopted to represent documents, where a document is represented as a multidimensional vector, while each dimension corresponds to a unique key term extracted from the documents. A common clustering technique is Self-Organizing Map (SOM), which is an unsupervised neural networks algorithm. Chen et al. [14] categorized a portion of the Internet documents with multilayered SOM to generate a hierarchical knowledge map system. Ong [9] employed an improved interface combining a 1D alphabetical hierarchical list and a 2D SOM island display to automatically generate a hierarchical NewsMap. To some extent, the knowledge map construction method based on automatic text classification is effective. But when applied in a real organizational environment, it suffers from many problems. First, the method relies much on the linguistic usage. Clearly, in organizational knowledge repository there are many kinds of files, such as drafts, videos, which have few words and consequently are not suitable for this kind of methods. Second, the design of taxonomy may not reflect business needs [6]. To address the above problems, an effective method for mapping enterprise knowledge is highly desirable.

2.2 Social Classification

Social classification refers to the collaborative way in which information can be organized on the web. It allows users to publicly add keywords to the shared

contents, as is totally different from the traditionally categorizing performed by an authority or authors. Users can not only categorize information for themselves, but also browse the information categorized by others. Keywords tagging is nothing new; the interesting thing is that when persons tagging in a public space, the collection of their keyword/value associations becomes a useful source of data in the aggregate [15]. Today, tagging is a widespread phenomenon popularized by applications such as social bookmarking (Del.icio.us) and social photo sharing (Flickr). It is discussed in some researches that tagging on the Internet has some limitations and weaknesses. For instance, ambiguity can emerge as users apply the same tag in different ways, while the lack of synonym control can lead to different tags being used for the same concept, precluding collocation [16].

We argue that social classification would be a fruitful way when it is applied in an organization. As the employees, especially those who work in the same or similar domain, share common business goals and have relatively similar business background, the problem in uncontrolled vocabulary will be solved to some extent. What's more, the social classification generated by employees will facilitate workplace democracy and the distribution of knowledge organization tasks among people actually using them, which is most helpful for an enterprise when there is nobody in the "librarian" role or there are too many unordered contents for few authorities to classify. So here we propose an improved social classification-based method to organize knowledge resources and construct enterprise knowledge map.

3   Outline of the Proposed Method

In this section, we describe knowledge map structure and the basic procedure of the proposed method for knowledge map construction.
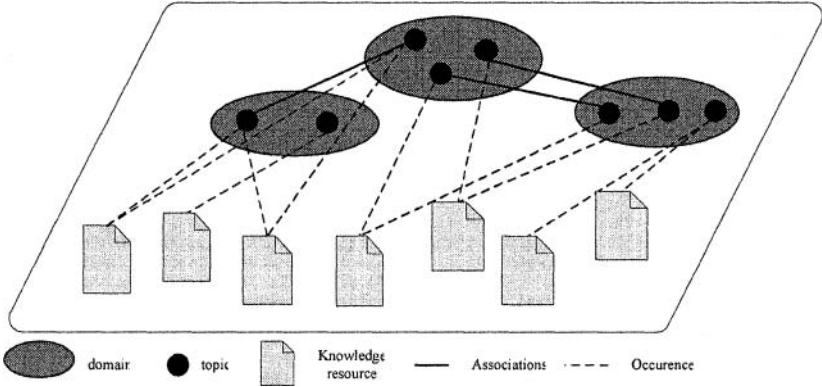
Fig. 1. Knowledge map structure

Fig. 1 shows the proposed knowledge map structure. To display knowledge and relationship, our proposed knowledge map is defined as sets of domain, topic, knowledge resource, relationships. For an enterprise-wide knowledge map, it is very difficult to map all kinds of knowledge into one picture. So here we define domain as the basic unit to organize knowledge and relationships. Domain is the context where a specific task, project or business will be carried out and also where knowledge is used, so domain-based knowledge mapping is more meaningful. Similar to topic map [13], our proposed knowledge map also consists of topics, which represent some concepts in a domain. Topics in different domains are related to each other by associations. A topic may also be related to knowledge resource by its occurrences.

```
                        ┌─────────────────┐
                        (      start       )
                        └─────────────────┘
                                 │
                                 ▼
            ┌──────────────────────────────────────────────┐
            │   Tag document with keywords by Individuals    │
            └──────────────────────────────────────────────┘
                                 │
                                 ▼
        ┌──────────────────────────────────────────────────────┐
        │ Calculate the information gain P(t) of each tag in a domain │
        └──────────────────────────────────────────────────────┘
                                 │
         no                      ▼                      yes
        ┌─────◇──────────────────────────────────◇─────┐
        │         P(t) >= IG. threshold                 │
        │   └──────────────────────────────────────┘    │
        ▼                                               ▼
┌──────────────────┐                        ┌──────────────────────┐
│  Delete the tag   │                        │ Select the tag as a topic │
└──────────────────┘                        └──────────────────────┘
        │                                               │
        └───────────────────────┬───────────────────────┘
                                 ▼
            ┌──────────────────────────────────────────────┐
            │      Form the knowledge map of a domain         │
            └──────────────────────────────────────────────┘
                                 │
                                 ▼
        ┌──────────────────────────────────────────────────────┐
        │ Calculate similarity sim(i,j) between topic i and topic j of │
        │               two different domains                    │
        └──────────────────────────────────────────────────────┘
                                 │
         no                      ▼                      yes
        ┌─────◇──────────────────────────────────◇─────┐
        │         sim(i,j) >= sim. threshold            │
        ▼                                               ▼
┌──────────────────────┐                    ┌──────────────────────┐
│ Disconnect the two tags │                    │  Connect these two tags  │
└──────────────────────┘                    └──────────────────────┘
        │                                               │
        └───────────────────────┬───────────────────────┘
                                 ▼
            ┌──────────────────────────────────────────────┐
            │      Form the organizational knowledge map      │
            └──────────────────────────────────────────────┘
                                 │
                                 ▼
                        ┌─────────────────┐
                        (       end        )
                        └─────────────────┘
```
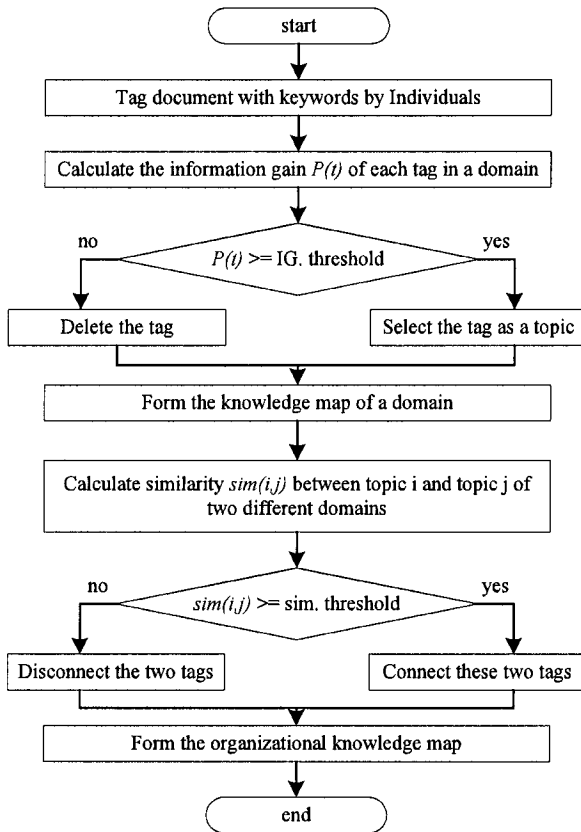
Fig. 2. The basic procedure of knowledge map construction

Fig. 2 shows the basic procedure of constructing the knowledge map. In Fig. 2, there are three major tasks associated with the knowledge map construction method: individual knowledge tagging, domain topic selection and inter-domain topic association analysis.

- Individual knowledge tagging is a process of social classification, which is the basis of the enterprise knowledge map construction. When individuals refer to some documents, which are helpful or important for his work, he will tag them. This is a process where knowledge works organize their knowledge.

- Domain topic selection is a process to construct a knowledge map for a team or department. When the members have classified the knowledge of interest, domain topics will be selected from the tags given by individual through an algorithm.

- Inter-domain topic association analysis is a process to construct the enterprise knowledge map. It carried out when two topics in different

domains are similar with each other to some extent.

Here we assume that all the classifications made by individuals are rational, i.e. they tag the documents according to their business needs. The relevant details will be discussed in section 4.

4   Knowledge Map Construction

In this section, the method to construct the enterprise knowledge map is described in detail. According to the framework mentioned in Section 3, the method is divided into three parts: individual knowledge tagging, domain topic selection and inter-domain topic association analysis. They correspond to knowledge classification on individual level, knowledge organization on team level and knowledge map construction on organizational level, respectively.

4.1   Individual Knowledge Tagging

Similar to the social classification, the method here also takes individual knowledge organization as a tagging process, in which users label some content they create or experience with one or more labels, or tags. However, tagging in a real enterprise environment is required to reflect the business needs, i.e. the tag given by a user should be comprehensive to others persons who work in the same domain.

The core idea of tagging must account for the full environment of social tagging. It is formalized as a three-place relation,

Tagging (object, tagger, tag).

Individual tagging is a process to form a personal knowledge map, which record every user's ideas about knowledge classification and his knowledge usage manners. It means more for the organization. The collection of tagging data will provide great insight into the status of allocating and applying knowledge and help to construct the enterprise knowledge map.

4.2   Domain Topic Selection

We have mentioned above that a domain provides a context where knowledge items and their relationship are displayed. When a specific task or project is carried out, a domain appears. As the project is going on, the tag space will consist of many unique terms words or phrases that is tagged on documents. It is highly desirable to reduce the tag space without sacrificing categorization accuracy.

According to Yang's study on feature selection techniques [17], information gain has relatively good performance. So here the domain topic is selected through analyzing domain members' tagging data and calculating the information gain of each tag. The tag t will be selected to be a topic when its information gain is greater than a predefined threshold α. The process can be expressed formally as follows.

Assume that the current tag is t, and the set of categories made by domain expert is $C = \{c_1, c_2, ... c_m\}$, the information gain of tag t is defined to be:

$$P(t) = -\sum_{i=1}^{m} P(c_i) \log P(c_i) + P(t)\sum_{i=1}^{m} P(c_i / t) \log P(c_i / t) + P(\bar{t})\sum_{i=1}^{m} P(c_i / \bar{t}) \log P(c_i / \bar{t})$$

If $P(t) \geq \alpha$, then the tag is selected as a topic of the domain.

4.3 Enterprise-Wide Knowledge Map Construction

In order to facilitate knowledge reuse among the organization scope, it is necessary to relate the topics of different domains by similarity, which is the process of the enterprise-wide knowledge map construction.

Topic (tag) here is represented in vectors

$$topic_x = \{w_{x,1}, w_{x,2}, \cdots w_{x,n}\}$$

Where each $w_{x,k}$ is a weight for document k for topic x. The weight of a document for a specific topic can be calculated referred to the definition of TF-IDF. We will not explain more about it here.

Similarity of two topics in different domain equals to cosine of the angle between them, i.e.

$$sim(topic_x, topic_y) = \cos(\theta) = \frac{topic_x \cdot topic_y}{\|topic_x\| \|topic_y\|} = \frac{\sum_k w_{x,k} \times w_{y,k}}{\sqrt{\sum_k w_{x,k}^2} \sqrt{\sum_k w_{y,k}^2}}$$

If the similarity between topic x and topic y is greater than a predefined threshold β, then we will build a connection between these two topics. Otherwise, there is no association between them

5   Knowledge Map System Implementation

Based on the proposed method, a prototype knowledge map system is being implemented in a knowledge management project for a Chinese manufacturing enterprise. The enterprise will be referred to as Company A due to confidentiality. Company A is a large state-owned aviation industrial limited company in the northeast of China. Its main business scope covers the design and development of airplanes. Design is a knowledge-intensive activity. After more than 50 years development, Company A have accumulated a large number of unordered knowledge materials, which leads to the ineffectiveness of knowledge searching and using, so it is an urgent task for company A to organize his knowledge resources and construct his own knowledge map.

The system employs Service-Oriented Architecture. Individual tagging, work center knowledge recommendation and organizational knowledge map construction are designed to be the three fundamental functions of the system.

Individual tagging is an important function in the system. It allows users to organize knowledge of interest during his work. When a user tags a document in the repository, the tagging action will be stored as a record and the tag will be saved as an attribute, as is shown is Fig. 3. Fig. 4 show the user interfaces of individual tagging.
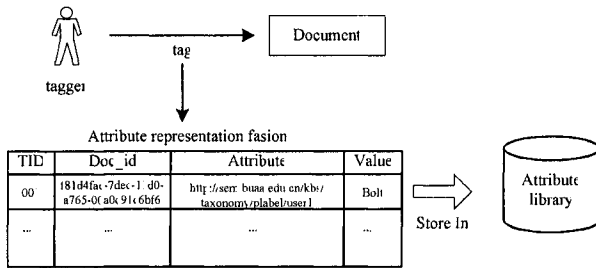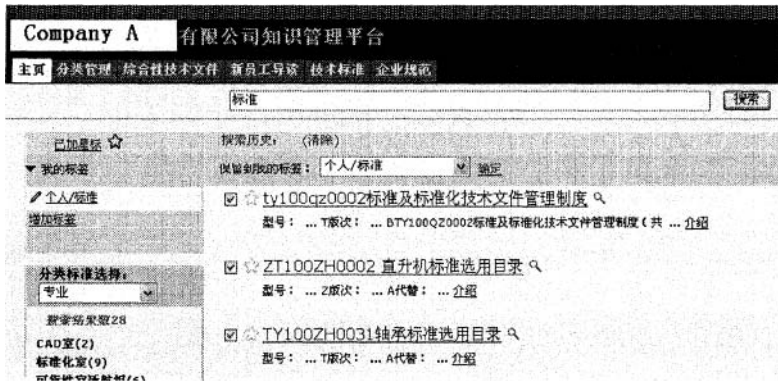

Fig. 3. Individual tagging

Fig. 4. The user interface

We design a virtual collaboration environment, work center, to manage the knowledge in a domain. Work center is a platform where members can share opinions about knowledge classification application. On one hand, every member's tagging data will be collected and analyzed, and a recommendation based on collaborative tagging will be given. On the other hand, an expert in this domain draws the domain knowledge map according to his expertise and experiences. The enterprise knowledge map is constructed when the topics in different domain is related to each other by similarity.

6   Conclusions

In this paper, we design a method for enterprise knowledge map construction based on social classification. Social classification is currently popular on the Internet. We argue collaborative tagging will provide some pragmatic benefits for enterprise knowledge organization. Unlike other methods for knowledge map construction, the proposed method develops the enterprise knowledge map as a multi-level system including knowledge tagging on individual level, topic selection on the team level and topic association on the organizational level. By virtue of the collaborative nature of social classification and the similarity of knowledge structure in a business domain, our method proposes that the knowledge structure in a business domain can be analyzed and integrated from individual knowledge tagging.   The enterprise knowledge map then is organized based on domain to display topics, knowledge resource, and their relationships.

     The project is still going on and further research is more challenging. With the implementation of the system, we can obtain more tagging data from users. It is necessary in further research to refine the algorithms in above method and validate them with the real data. Tagging and work center (domain) are both meaningful settings for knowledge organization in the enterprises. Another crucial problem is what kind of measure we can take to drive more workers to employ them.

Acknowledgements

Reference

1. M. Yogesh, Knowledge Management and New Organization Forms: a Framework for Business Model Innovation, *Information Resource Management Journal* **13**(1), 5-14 (2000).
2. R.L. Chase, The Knowledge-Based Organization: An International Survey, *The Journal of Knowledge Management* **1**(1), 38-49 (1997).
3. L. Argote, S.L. Beckman, and D. Epple, The Persistence and Transfer of Learning in Industrial Settings, *Management Science* **36**(2), 140-154 (1990).
4. E.D. Darr, L. Argote, and D. Epple. The Acquisition, Transfer and Depreciation of Knowledge in Service Organizations: Productivity in Franchises, *Management Science* **41**(11), 1750-1762 (1995).
5. M.J. Eppler, Making Knowledge Visible Through Intranet Knowledge Maps: Concepts, Elements, Cases, in Proceedings of HICSS-2001, the 34th Hawaii International Conference on System Sciences (3-6 Jan. 2001).
6. A.D. Marwick, Knowledge Management Technology, *IBM Systems Journal* **40**(4), 814-830 (2001).

7. T. Davenport and L. Prusak, *Working Knowledge: How Organizations Manage What They Know* (Harvard Business School Press, Boston, 1998)

8. E.F. Vail III, Knowledge Mapping: Getting Started With Knowledge Management, *Information Systems Management* **16**(4), 16-23 (1999).

9. T. Ong, H. Chen, W. Sung, and B. Zhu, Newsmap: A Knowledge Map For Online News, *Decision Support Systems* **39**(4), 583-597 (2005).

10. K.M. Drabenstott and M.S. Weller, The Exact-Display Approach for Online Catalog Subject Searching, *Information Processing and Management* **32**(6), 719-745 (1996).

11. M. Massicotte, Improved Browsable Displays for Online Subject Access, *Information Technology and Library* **7**(4), 373-380 (1988).

12. J.D. Novak and D.B. Gowin, *Learning How to Learn* (Cambridge University Press, New York, 1984).

13. S. Pepper, The TAO of Topic Maps: finding the way in the age of infoglut (April 2002); http://www.ontopia.net/topicmaps/materials/tao.html

14. H. Chen, C. Schuffels, and R. Owig, Internet Categorization and Search: A Machine Learning Approach, *Journal of Visual Communications and Image Representation Science* **7**(1), 88–102 (1996).

15. T. Gruber, Ontology of Folksonomy: A Mash-up of Apples and Oranges, The First On-Line Conference on Metadata and Semantics Research (November 2005); http://tomgruber.org/writing/ontology-of-folksonomy.htm.

16. A. Mathes, Folksonomies-Cooperative Classification and Communication through Shared Metadata (December 2004); http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html.

17. Y. Yang, and J.O. Pedersen, A Comparative study on Feature Selection in Text Categorization, in Proceedings of ICML-97, 14th International Conference on Machine Learning, pp. 412-420 (1997).