

## The Description of Blog Resource Based On Ontology

Zhijuan Yang

School of Economy and Management, Xidian University, Xi'an 710126, China.

zhijuan\_yang@126.com

### Abstract

Nowadays, people are lack of fast access to the content of blogs, which makes the browsing and using blogs is less frequent than those of micro-blog. The semantic retrieval of blog content is realized from the key basic level of resource description, which is easy to obtain semantic information and knowledge contained in Bowen in knowledge community. In this paper, ontology is used to construct an ontology based blog resource description model, which systematically represents the external features of blogs, the content features of blogs, and the semantic relations between these features. Taking an Bowen as an example, the experiment verifies the feasibility of the model and provides theoretical support for accurate and efficient Bowen knowledge retrieval.

### Keywords

Blog, ontology, text resources description.

### 1. Introduction

The Web2.0 era is the Internet age of data explosion, where people acquire scientific knowledge from online communities, such as blogs, wikis and zhihu. Blog is not only an open platform for human communication, communication, emotional expression and knowledge exchange, but also as a storage carrier for data, information and knowledge. It Not only contains a circle of communication between different users, and generates a large number of symbols, pictures, words, sentences, paragraphs or other character arrangement content form to express or explain human needs all kinds of knowledge[1]. The traditional way of accessing blog resources is to obtain the semantic information and knowledge required by users by browsing the content of the blog word by word. In the era of big data background, blog resources have massive, heterogeneous, real-time updating features, it is difficult for users to quickly obtain the required knowledge from the mass of blog resources. Therefore, how to quickly search and access from the knowledge level is urgent for solving the problem of the knowledge contained in Bowen.

The structured description of blog resources is the key link to solve the problem of Inefficient access to Bowen knowledge. According to the characteristics of blog resources, the scientific and rational description method of resources, can describe the unified structure of important information and knowledge in blog resources, so as to further realize knowledge storage and orderly organization for the massive, heterogeneous and dynamic text resources . The text in the text resources of blog resources contains the knowledge points required by the user knowledge, but numerous data and information elements of text resources used to be described, integrated and organized by the Dublin core metadata, such as the title, creator, subject, description, publisher, participant, date, type, format identifier, origin, language, relationship, coverage, copyright and so on[2]. These metadata can only describe the external features of blog text resources, but can't describe content feature of knowledge points in the blog text resources. From the knowledge level, the fine description of knowledge resources in Bowen's content is easy to discover and extract relevant knowledge points in Bowen, and realize the organization, storage, reasoning and innovation of knowledge.

On the basis of ontology theory, this paper constructs an ontology based blog resource description model. This model uses ontology knowledge representation to describe external features of blog resources and semantic features of blog content. The ontology conceptual hierarchy and logical

reasoning ability can describe and present the semantic structure and relationship of the concepts in blog resources. The experiment uses a blog post to verify the feasibility of the blog resource description model.

## 2. Literature Review

### 2.1 Description Method of blog resources

Blog resources not only include user characteristics, such as bloggers, user comments, user tags, etc., but also contain a large number of blog content and other content, such as blog topics, event content. There are many ways to describe the characteristics of blog users, mainly divided into the following three: ①Based on self-organization theory, the construction of Bowen correlation model to describe the overall network and small world effect contained in the blog post[3]; ②Usage of statistical methods to analysis semantic information, such as user tags, blog posts frequency, blog hot topic[4]. ③Usage of natural language methods and data mining methods to analysis of user behavior and Bowen content and minesocial network relationships and information exchanging rules contained in blog[5].

The description of content features of bowen is based on the theory of text resource description. Common methods of text resource description include based on ontology theory, based on HNC theory[6], based on knowledge element theory, graph theory, based on statistical theory and method to construct a Model that covers a variety of semantic features of the resource, such as the construction of the conceptual hierarchy model [7], the multivariate ontology hierarchy model[8], the semantic mapping of the concept and topic[9], the semantic link network data model[10][11], the semantic model of Articles (SMA)[12], Ontology knowledge base and so on. The above text resource description method can be used to discoverknowledge link network in knowledge community[13], the related documents in collection resources[14], the integration of knowledge in the field of mechanical manufacturing[15], the organization and presentation of scientific knowledge in academic literature[16], and so on. For the description of blog content, classify blog information, the subject and the content of the word characteristics into two levels structure[17][18], or describe bloggers, publishing time, log links, tags and other basic information[19], but blog's built-in search system is based on a simple keyword matching and lack of knowledge of the expression, processing and understanding capability, which is difficult to achieve the semantic retrieval for blog resources[20].

The above theories and methods can be used to describe the user resources or text resources in various fields, but can't comprehensively describe the finer content characteristics of blog resources. Using the ontology theory can make all kinds of fine-grained knowledge hierarchical, standardized to improve the blog storage and retrieval mechanism.

### 2.2 Ontology-based knowledge representation

Ontology is a clear formal specification of shared conceptual models with shared, explicit, conceptual and formalized features[21].The ontology has a good conceptual hierarchy structure and support for logical reasoning, which can be used to describe the complex semantic relations between concepts and concepts and also be used in semantic rules reasoning, knowledge discovery, knowledge innovation and so on[22].In order to realize the formal definition of ontology, it is necessary to form the formal language of description ontology, including XML, RDF, RDFS and OWL[23]. By constructing a knowledge ontology model in a certain field, we can describe the semantic rules between information and knowledge in this field, and provide sufficient support for knowledge retrieval, which lay a good foundation for knowledge reasoning[24][25].

Ontology-based knowledge representation has a strong hierarchical and logical reasoning ability, not only describes the hierarchical structure of knowledge in text resources, but also derives the implicit association between knowledge based on the explicit semantic association between knowledge, and

realize knowledge discovery and innovation. At present, people according to different application requirements, has built a lot of specific field ontology. The mature application ontology is frequently appeared in medicine field[26][27], emergency emergency command[28] and other fields, but less in the field of blog resource semantic description.

Therefore, this paper will use the ontology theory to construct a ontology-based blog resource description model to comprehensively describe the semantic features of blog resources, and provide an effective resource description method to realize the semantic retrieval of blog resources.

### 3. Blog resource ontology description model

Combined with ontology knowledge representation, this section builds a ontology-based blog resource description model, that is, blog resource ontology. Firstly, This model describes conceptual hierarchy structure of blog resource ontology, and make the semantic relations between concept classes more specialized and systematic; secondly, describe the conceptual attribute that is used to fill the ontology framework and represent more specific knowledge objective; finally, The axiom and examples of resource ontology is used to describe the semantic relations between knowledge objects in blog resources. From the fine-grained level, a complete resource description model is presented to describe the characteristics of blog text resources.

#### 3.1 The Conceptual Hierarchy of Blog Resource Ontology

The blog resource description model describes the blog resource characteristics from three levels: the resource layer, the information layer, and the text layer. The details are as follows:①In the resource layer, the resource description object is a resource file, which describes the resource identifier, resource storage address, resource storage date and resource text type of resource file. There are two kinds of performance text, namely, comment class and blog class. ②In the information layer, the resource description object is the resource text, whose features can be subdivided into personnel, website, text format, grade classification and reference link, personnel, including viewers, collectors, represent, critics and bloggers; Web site can be divided into website address and name; text format includes graphics, formulas, text three categories; grade classification includes blog level, attention, honor badge, and attention includes blog visitors and popular bloggers; Reference links include Anchor tags and link addresses.③In the knowledge layer, the resource description object is the knowledge in blog resource, and the knowledge resource is described from the four perspectives that is knowledge navigation, knowledge content, knowledge style and domain theme. The knowledge navigation includes the link page and identification number; the knowledge content includes name, feature phrases and the original text description; the knowledge text includes narrative text, description style of text, discussion style of text, and application style of text. Narrative style of text also includes narrative event, narrative person, landscape written, shape of object; discussion style of text includes the commentary type, the essay type and the political type of knowledge; the description style of text includes the entity thing, the fact explanation and the science stories; the application style of text is divided into eight categories that includes letters, notes, documents, propaganda, notices, rituals, contracts, and biography; domain theme class is constituted by classification and theme class, which can also uses the "Chinese classification theme list" to refine the theme class.

#### 3.2 Conceptual Attributes of blog resource ontology

The concept of attribute blog resources in the ontology includes data attributes and object attributes, in which the data attribute describes the attributes of concept class, including personnel attributes, Explanatory genre attributes, narrative style of text attributes and so on; the attributes of the personnel includes name, date of birth, gender; description type of text's attributes includes manufacture, function, manufacturing methods and status; narrative style of text's attributes includes time, place, people, events cause, events pass, event result.

Object attributes describe the semantic relations between concept classes, which can be divided into the following two categories:①The semantic relations of the same level, that is, the semantic relations between the conceptual classes at the same level, can be divided into explicit semantic

relations and implicit semantic relations. Among them, the dominant semantic relations include the relationship between time and order, the relationship between the upper and lower, the Reciprocal relationship, the relationship of order, the relationship of cooperation, the relationship of decomposition and the relationship of reference; implicit semantic relations includes the relationship of extrapolation, implied relationship, positioning relationship, cross mappings relationship, and so on. ②The semantic relations of different layers, that is, the semantic relations between the conceptual parent and the concept subclass, includes the same relationship, the whole part relationship, the similarity relationship, the attribute relationship, the instance relationship, the classification relationship and so on.

In the blog resource ontology, see Fig. 1, shows the above part of the semantic relation, in which the resource storage time in the resource layer has the time semantic relation; the viewer in the information layer contains the collector, the repacker, the commenter, which is including relationship; Knowledge link page in the knowledge layer has a corresponding relationship with the link address in the reference link in the information layer, which is related relationship; the link is published in the web content of the website, that is, the publication relationship.

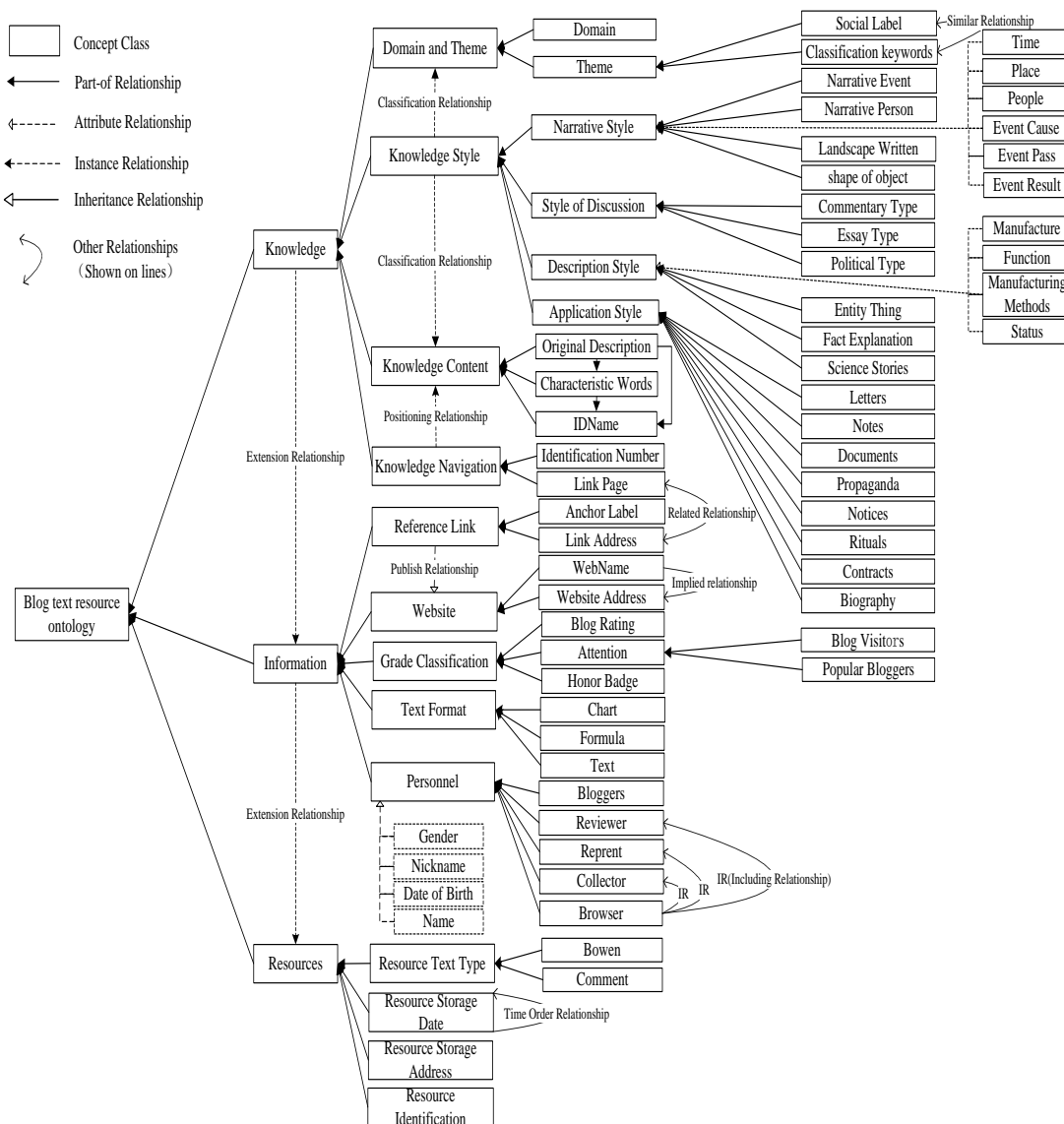


Fig. 1 Blog resource ontology

### 3.3 The axiom and instance of blog resource ontology

The axiom and examples of the blog resource description model are described as follows: The axiom of the ontology describes the forever truth of the domain knowledge, which can describe the rules of

semantic relations and the node attributes. The axiom of the semantic relations rule class can be described as that the relationship between reference relationship and the cited relationship is reciprocal; the inheritance relationship and the extension relationship is expansive and unidirectional; the cooperative relationship and the same relationship are bidirectional. The axiom of semantic relations between concept node can be defined that the domain and range of the time relationship are Date type; the domain and range of the whole-part relationship may be concept class or instance; and the reference does not intersect with cited objects in personnel class. The instance in the ontology is an instantiation process of one or more abstract concepts, which includes certain attribute values, and only a few important instances are generally considered in the ontology construction process

#### 4. Case Study

Experiment selects a Sina blog of popular king of Han Han, and extract a narrative type of blog that is named of "Birth of That tea sister's "ride the wind and waves", as this experimental data source.

First of all, use the Protégé tool to show the concept classes, the concept of attributes, axiom of blog resource ontology. This section can only exhibit a part of the blog resource ontology shown in Fig. 2.

Because of semantic features in blog resource ontology are quite comprehensive, we only shows a part of semantic features that is related to the concept of "knowledge style" in the ontology model. Extract the knowledge points of that hanhan's blog, and use blog resource ontology to represent relevant knowledge points in that blog. The results of instance is shown in Fig. 3.



Fig. 2 Blog resource ontology (a part of the ontology)

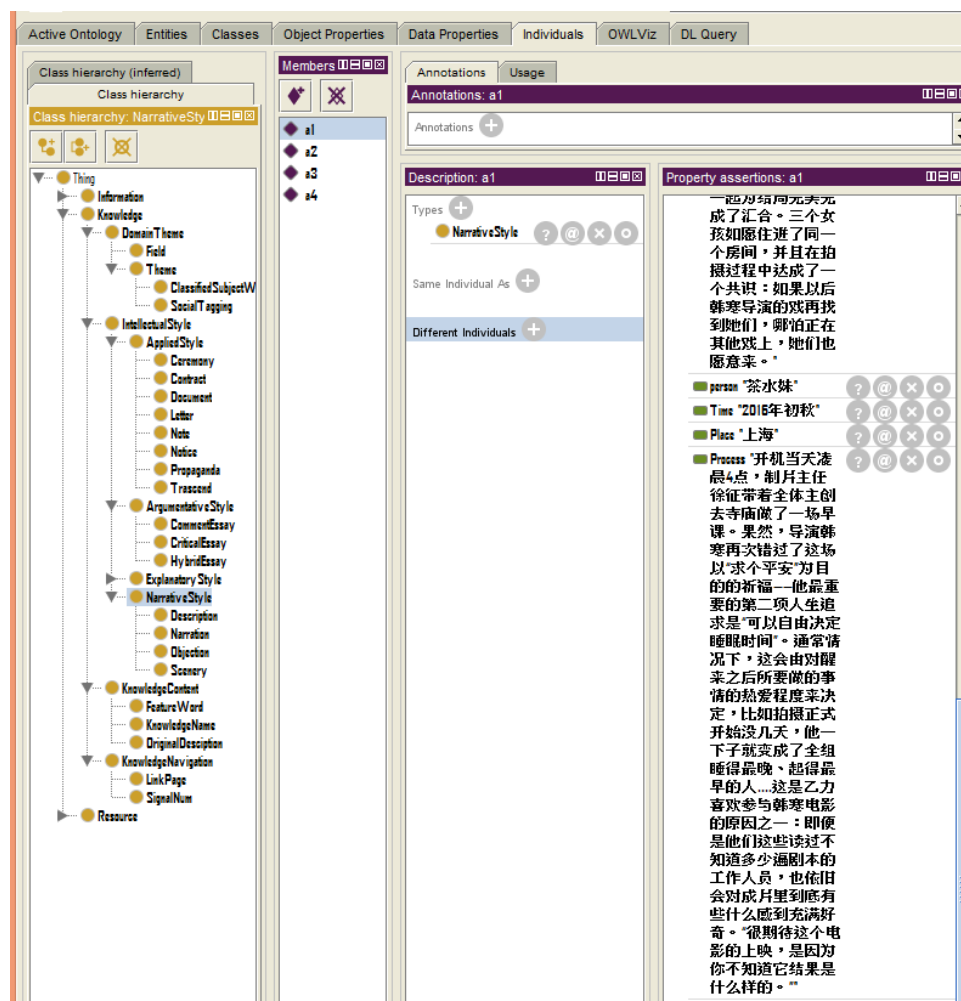


Fig. 3 Instantiation of blog resource ontology

## 5. Conclusion

Based on ontology theory, this paper constructs a ontology-based blog resource description model, which integrates and expresses the fine granular knowledge in blog resources, and provides some theoretical support for the precise semantic retrieval of blog resources. However, experiment only shows the instantiation process of parts of concepts in "knowledge ontology" and fails to show the instantiation of throughout ontology. In addition, only using a case is not enough to verify the rationality of the ontology model, Thus we will continue to modify the blog resource ontology and further study the evaluation Method of blog resource ontology.

## References

- [1] Duan R T: Research on standardization of XML integration standard of Chinese electronic document knowledge organization,(Shanghai Jiao Tong University press, 2012).
- [2] S Bird: Natural Language Processing with Python(Posts and Telecommunications Press, 2014).
- [3] Lin-Hong L I, Rong-Rong L I. Studies on Self-organization Behavior of Sina Micro-blog Social Network. Statistics & Information Forum, Vol.28 (2013) No.1, p.88-94.
- [4] LEE D H, SCHLEYER T. Social tagging is no substitute for controlled indexing:a comparison of medical subject headings and Cite ULike tags assigned to 231,388 papers. Journal of the American Society for Information Science and Technology, Vol. 63(2012) No.9, p.1747-1757.
- [5] Hu Y. Informal Information Flow Mechanism Based on Timing of Micro-blog——Example of Sina Moco-blog. Document,information & Knowledge,(2011) No.4, p.111-117.
- [6] Li S, Wang Y. Retrieval of Periodical Literature Knowledge Element Based on Hierarchical Network of Concepts Theory. Journal of Intelligence, (2013) No.9, p.190-194.

- [7] Li J F, Lv X Q, Li Z, et al. Deriving Concept Semantic Hierarchy of Ontology in Patents. Journal of the China Society for Scientific and Technical Information, (2014) No.9, p.986-993.
- [8] Ma S, Tian L. Ontology-based semantic retrieval for mechanical design knowledge. International Journal of Computer Integrated Manufacturing, Vol. 28(2015) No.2, p.226-238.
- [9] Morales L P, Esteban A D. Using Semantic Graphs and Word Sense Disambiguation Techniques to Improve Text Summarization. Procesamiento Del Lenguaje Natural Revista, (2011) , p.97-105.
- [10] Xu Z, Liu Y, Mei L, et al. Semantic based representing and organizing surveillance big data using video structural description technology[J]. Journal of Systems and Software, (2015) No.102, p.217-225.
- [11] Hai Z. Communities and Emerging Semantics in Semantic Link Network: Discovery and Learning[J]. Knowledge & Data Engineering IEEE Transactions on, Vol.21(2009) No.6, p.785-799.
- [12] Marcondes C H. Representing and organizing scientific knowledge in biomedical articles with Semantic Web technologies. (2014).
- [13] Hai Z. Communities and Emerging Semantics in Semantic Link Network: Discovery and Learning. IEEE Transactions on Knowledge & Data Engineering, Vol.21 (2009) No.6, p.785-799.
- [14] Qiu J P, Xu C. Research on the Library Resource Ontology Semantic Similarity Measuring. The Journal of the Library Science in Jiangxi, Vol.45 (2015) No.3, p.1-7.
- [15] Bernard A, Boucher X, Vajna S, et al. International Journal of Compute Integrated Manufacturing. International Journal of Computer Integrated Manufacturing, Vol.23 (2010) No.2, p.146-154.
- [16] Marcondes C H. Representing and organizing scientific knowledge in biomedical articles with Semantic Web technologies. (2014).
- [17] Zhao Y. The Design of a Semantic Blog System Based on Ontology[J]. Journal of Intelligence, Vol.28 (2009) No.1, p.43-46.
- [18] Zheng M Y. Research of an Ontology-Based Chinese Blog Automatic Secondary Classification. Information Science, Vol.34 (2016) No.2,p.87-90.
- [19] Tang Y F, Lin J L. Library and Information Blog Ontology Construction and Implementation of Semantic Search System. Journal of Library & Information Sciences in Agriculture, Vol25 (2016) No.8, p.40-44.
- [20] Kong C. Research on the Application of Ontology Technology Used in Semantic Blog.Computer & Digital Engineering, (2011).
- [21] Studer R, Benjamins V R, Fensel D. Knowledge engineering: Principles and methods. Data & Knowledge Engineering, Vol.25 (1998) No.1-2, p.161-197.
- [22] Jiao Y Y, Wen Y K, Lu W. New theory of information retrieval(Wuhanuniversity press, 2008).
- [23] Wu G B. Ontology-based information retrieval service research of teresources( East China Jiaotong University, 2013).
- [24] Xu H, Musicant O. Design and Implementation for Ontology Modeling of Design Knowledge Based on UML Class Diagram. Vol.14 (2016) No.3A, p.326.
- [25] Liu J. Ontology-based Semantic Reasoning of UML Class Diagram. Computer Applications & Software,(2011).
- [26] Brown W. Ontology-based Semantic Harmonization of HIV-associated Common Data Elements for Integration of Diverse HIV Research Datasets(Columbia University, 2016).
- [27] Edwards B I, Muniru I O, Cheek A D. Robots to the Rescue: A Review of Studies on Differential Medical Diagnosis Employing Ontology-Based Chat BotTechnology. (2016).
- [28] Qiu J N, Wang X H. Model and method of emergency knowledge management(Science Press, 2016).