

## SEMANTIC ANNOTATION AND STORAGE FOR TOURISM INFORMATION

Yezheng Fan

China Academy for Tourism Talent Development, Beijing International Studies University,  
Beijing, China, 100024.

Email: fanyezheng@bisu.edu.cn

### ABSTRACT

In recent years, semantic retrieval for web information becomes a hot motivation of research. Semantic annotation and storage for web information is very important as the basement of semantic retrieval. In this paper, we compare the two semantic annotation methods: specialist annotation and social annotation. We also compare the two semantic storage strategy: database and file system and we propose the function of semantic annotation and storage system.

**Keywords:** Semantic annotation, Semantic retrieve, Ontology, Tourism information.

### 1. INTRODUCTION

With the information explosion in the web and the explosion of people express their queries, search engine need to understand the meaning of information and query to improve the accuracy of information retrieval [1]. Semantic search which based on semantic web is a brand new search which the semantic annotation of resources is the foundation of the semantic computing. Such new search engine is semantic search engine which can diminish the gap between what the user really require and what the search engine can provide [2].

The key technologies of semantic search include domain ontology, semantic annotation, storage strategy, inference mechanism and semantic ranking etc. Domain ontology often includes domain concept, domain instance, relations between domain concepts and some axiom and deduction in special domain [3]. Thus domain ontology can describe the semantic of information in special domain. These ontologies are the basic element of semantic search [4].

Semantic annotation is the basement of semantic search. It annotates information with ontology elements. Annotation tools are often explored to fulfill auto or half auto annotation for information.

Storage strategy is the key technology for semantic search because it will affect the efficiency and usability of the whole search system. Different kinds of concepts and instances should adopt different storage strategy to improve efficiency and usability. There are two main types of storage strategy: file system and database system. File systems strategy often store non-structure data. And database often store structure data.

Inference Mechanism is the procedure that getting semantic results through defined relations in ontology. Semantic inference is often built on description logic that is the logic base of semantic web.

There are only two results for traditional database or keyword retrieval. They are “yes” or “no” (1 or 0). So there is no sort of problem, but for semantic retrieve, there may be many results which relevant to the query. So semantic ranking is important for semantic retrieve. We can refer to some sorting methods of traditional search engines, adding more semantic factors to establish a mathematical model to calculate the degree of semantic relations [5].

### 2. ONTOLOGY OF TOURISM INFORMATION

Ontology defines the semantic relations between domain concepts. The main relation of these concepts is subsume relation. So the domain classification category is often used as domain ontology in application systems. About the category of tourism information, there exist many classification catalog. Take China as an example, there are at least two kinds of catalog. One is the industry catalog which is show in Fig.2, the other is the academy catalog which includes at least three different catalog: the Chinese library classification, CNKI (Chinese National Knowledge Infrastructure) classification and RUC (Renmin University of China) database classification. Because most websites use industry classification to exhibition their information, we take the industry classification in Figure.1 as domain concepts so that the tourism knowledge may be shared intensively according this classification.

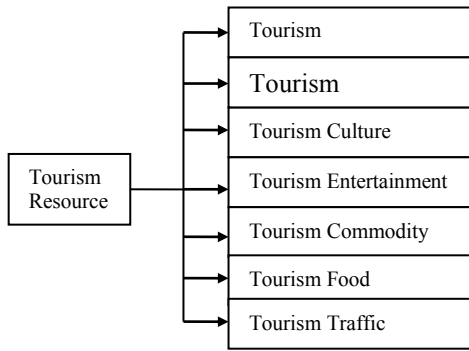


Figure 1. Tourism resources ontology

### 3. SEMANTIC ANNOTATION AND STORAGE

#### 3.1 Two types of annotation

Semantic search which based on semantic web is a brand new search in which the semantic annotation of resources is the foundation of the semantic computing. There are two main kinds of annotation methods: specialist annotation and social annotation [6].

Specialist annotation often related with ontology. Ontology is a semantic model which is the popular research point in recent years, which brings more complex relationship to dictionary and taxonomy [7]. Ontology is based on description logic, which supports logic reasoning, thus can more clearly, accurately express semantic and knowledge. As a philosophical concept, ontology is introduced by Neches in the field of artificial intelligence research. Its definition is that giving basic terms which constitute the vocabulary in related field, and make the definition of rule of these vocabulary extension by using the rule consist of these terms and relation. Taxonomy or directory (classified catalogue) can be thought as a kind of simple ontology.

Social annotation [8] is the newest word with the development of the Web, which is consist of folk and taxonomy. Semantic mark, which is based on folksonomy, can be added freely by the users of the whole society. For example, users can add label to any pictures in photo browser and users can add label to bookmark in delicious bookmarks. Folksonomy and social annotation can be in place of the traditional keywords, and can calculate the semantic similarities using these annotations, thus to improve the related sort of Web pages [9].

Table 1 compares the advantage and disadvantage of two kinds of annotation method in different aspects.

Table 1. Comparison of two kinds of semantic annotation

	specialist annotation	social annotation
verifiable	better	poor
compatibility	poor	general
usability	general	better
unambiguous	better	poor

#### 3.2 Annotation system

The annotation system should include two parts. One part is responsible for building ontology. The other part is responsible for annotation instances (or resources).

Ontology building part includes the following functions:

(1) Define domain concept. Define the concepts name and their properties.

(2) Define relationships between domain concepts. The main relation between concepts is subsume relation. Some other relations such as similar, sequence, part-of etc. can also be defined.

(3) The storage strategy for concepts, their properties and relations, including concept there are two main storage strategies: database and file system. Compare these two strategies, database model is mature, and has many approachable application API for its implementation, not only in construction and storage is low and limited. Files such as RDF/OWL files with more serviceable semantic meanings can store and search concepts and instances by means of semantics [3], while it's retrieval efficiency is faraway to meet demand.

(4) After defining concepts, their properties and relations, the whole domain concept architecture has been achieved and we can draw the main structure in a single image. The structure can be seen as the knowledge schema, whereas, what people concern more are instances belonging to those concepts, and how they inherit the properties and relationships defined in the concept. Thus, in addition to store those schemas and draw the schema structure, we should found the related data structure or schema information for following instance construction.

Instances annotation part includes the following function:

Semantic annotation for concept discussed above is the basis of semantic annotation for instance. As the result of the former structure, each concept has its own integrated schema and storage model. With those schemas, we can define and manage related instances. Coming to the instances conservation, there are also two strategies, namely, database storage or semantic web storage such as RDF. Using database, the related data tables are constructed dynamically according to the former concepts, so each concept has a table, and sometimes we also need tables to store the relationships. For example, we have cave painting table to store papers, dynasty table to store dynasty and cave painting to dynasty table to store relationships between cave painting and dynasty; while using RDF files directly, each concept has a RDF schema, and the defined instance ought to be coincident with the schema strictly. The information to be conserved about concept is less than that of instance, while the concept has more semantic relationships to be stored and retrieved. As we know, database is mature enough for information system, not only in scientific research but also in industrial application, while the semantic web storage models, such as RDF, OWL, are excellent in semantic modeling, because of their rich semantic relationship description [10].

So the instance annotation function consists of a set of sub-functions. Each sub-function provides management interfaces for each concept to manage its own instances, including instance definition, modification and deleting.

## 4. SEMANTIC SEARCH PORTAL

With the two annotation methods mentioned above, we have constructed and stored domain concepts and relevant instances as a whole integrated resource base.

The two methods can be regarded as a kind of semantic-driven resource management model. Whereas, the management model is just a means, the goal is resource retrieval by means of semantic model. Therefore, we propose semantic portal for those resources retrieval in semantic way, which can be ascribed to abundant semantic information, including concepts, instances and relationships, built in the two semantic annotation tools.

Different from traditional search engine like Google, the semantic portal gives the search results in semantic mechanism, in which, the query is not just non-sense keyword but with some useful semantic meanings. Meanwhile the retrieved resources are not just in non-sense results list, but organized by semantic structures [11].

Even we can get the search results as a single whole semantic picture, which gives searchers a schematic retrieval resource structure with intuitive but abundant semantic relationships. Compared with search interface in traditional resource or information management system, interfaces in the semantic portal are for both instances and concepts, and even for three kinds of relationships as mentioned before. All of those results retrieved are attained and organized by semantic strategies. Therefore, we call it semantic search portals.

### 4.1 Semantic search interface for concept

This portal provides the interface for searching domain concepts. In the authoring tool for domain concept, we have defined knowledge schema of domain ontology as a whole structure, which includes domain concepts and relationships between those concepts [12].

Given a domain concept, the portal will return the concept's schema, including basic information of the concept, the data-type properties and object properties. Taken "researcher" as example, the portal will give us the schema guideline about how to depict "researcher" in semantic way, including the data-type properties, such as name, degree, homepage, etc., and object properties, also known as relationships with other concepts, such as research topics, supervised students, even super-class "Person" etc.

### 4.2 Semantic search interface for instance

In the search portal, to achieve much more semantic strategies, we use front-end semantic methodologies. As strategies discussed in Section 3 for excavation of semantic meanings, we only adopt semantic annotation for search query, that is, when submitting a query with keywords, user can also denote it some concept (or object) [13].

In research community, the domain concepts are limited, and in our prototype system, there are limit concepts, those concepts are fairly stable and the relationships are moderately clear. Therefore, people pay more attention to their relevant instances. Different from traditional search engine, the portal provides the retrieved resources not in the non-sense list without any structure information for each resource, but in the objects or resources list, in which each object can be seen

as a resource entity with much semantic meanings, not only with structure depiction but also with some other relative resources by semantic association.

### 4.3 Semantic search interface for relationship

In addition to single concept or instance, we also need to discover relationships between those domain resources. As a simple mathematic model, the relationship between two kinds of entities can be classified into three categories. Thus, in our semantic search portal for relationship, we have three sub-interfaces, namely, relationship between concept and concept, relationship between instance and instance and relationship between concept and instance. In the following, we will discuss those three kinds of relationship retrieval.

(1) Relationship between concept and concept. Given two concepts, it is useful and interesting to discover the semantic association between them. Taking "researcher" and "research team" as example, we may find their semantic relationship "MemberOf"; although we haven't defined the "MemberOf" relationship during the authoring, but as "Researcher" is subclass of "Person" and "Person" has relationship "MemberOf" with "Research Team", we can safely conclude the relationship "MemberOf" between "Researcher" and "Research Team". Therefore, relationship can be inferred, according to sub-class properties, and in the future, by absorbing the semantic links, we may get much more inference between those concepts.

(2) Relationship between concept and instance. Given one concept and one instance, we need to discovery the semantic association between them. For example, when instance "YZ. Fan" and concept "Person" are given, we may attain semantic relationship "YZ. Fan" is an Instance of Person" retrieved from our implementation system. Although we haven't defined the instance of "YZ. Fan" in the authoring tool for concept "Person", it has been defined as an instance of "Teacher". As defined, "Teacher" is a subclass of "Person". Consequently, according to the following reasoning rule, we may safely draw a relationship "YZ. Fan is an instance of a person" as a conclusion. More powerful and efficient reasoning rules can be obtained or adopted from semantic link.

(3) Relationship between instance and instance. The relationship between instances is interesting and abundant. Given two instances, we need to discover the semantic association between the instances. With the increase of the number of instances, those relationships may grow at exponential rate. As a result, in our current implementation, semantic relationship retrieval between instances is limited, and we haven't proposed an effective strategy for ranking those semantic relationships yet.

## 5. CONCLUSIONS

Semantics can deeply impact the search with its widespread availability, and it may push us into a new search era. Based on the ontology, this paper proposes a kind of semantic annotation system for tourism domain. It includes ontology building and instances annotation. Upon the semantic annotation, semantic search can be accomplished [14].

Our ongoing work is using the tourism ontology to annotate the pictures in our tourism information system. Aided by these semantic annotations, we may build semantic picture retrieve system.

## ACKNOWLEDGMENTS

This work is supported by the scientific planning project of Beijing municipal commission of education (SQKM201410031001), the National Natural Science Foundation of China (No. 61300132).

## REFERENCES

1. R. Guh, R. McCool, E. Miller, Semantic Search, WWW 2003, *Proc. of the 12th International Conference on World Wide Web*, 700-709. ACM Press, 2003.
2. Calsavara, G. Schmidt, Semantic Search Engines, ISSADS 2004, 145-157, Springer 3061, 2004. DOI: [10.1007/978-3-540-25958-9\\_14](https://doi.org/10.1007/978-3-540-25958-9_14).
3. T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, *Scientific American*, May 2001. <http://dx.doi.org/10.1038/scientificamerican0501-34>
4. S. Staab, R. Studer, Handbook on Ontologies, Springer, January 2004. DOI: [10.1007/978-3-540-24750-0](https://doi.org/10.1007/978-3-540-24750-0).
5. C. Rocha, D. Schwabe, Poggi de Aragao M., A Hybrid Approach for Searching in the Semantic Web, *WWW 2004*. DOI: [10.1145/988672.988723](https://doi.org/10.1145/988672.988723).
6. Golder S. A., Huberman B. A., The Structure of Collaborative Tagging Systems, *Arxiv Preprint*, 2005.
7. Paul Buitelaar, Philipp Cimiano, Stefania Racioppa, Melanie Siegel, Ontology-based Information Extraction with SOBA. In: *Proc. of LREC 06*, pp. 2321-2324, 2006.
8. J. Hendler, T. Berners-Lee, From the Semantic Web to Social Machines: A Research Challenge for AI on the World Wide Web, *Artificial Intelligence*, Volume 174, Issue 2, February 2010, pp. 156-161.
9. Yue Lu, Malu Castellanos, Umeshwar Dayal, Cheng Xiang Zai, Automatic Construction of a Context-Aware Sentiment Lexicon: An Optimization Approach, *WWW 2011*, Hyderabad, India.
10. A. Calsavara, G. Schmidt, Semantic Search Engines, *ISSADS 2004*, 145-157, Springer 3061, 2004. DOI: [10.1007/978-3-540-25958-9\\_14](https://doi.org/10.1007/978-3-540-25958-9_14).
11. A. Cockburn, Writing Effective Use Cases, Pearson Education, 2002.
12. A. Gomez-Perez, O. Corcho, M. Fernandez-Lopez, Ontological Engineering: with Examples from the Areas of Knowledge Management, *E-commerce and the Semantic Web*, Springer, 2003.
13. C. Rocha, D. Schwabe, Poggi de Aragao M., A Hybrid Approach for Searching in the Semantic Web, *WWW 2004*. DOI: [10.1145/988672.988723](https://doi.org/10.1145/988672.988723).
14. Y. Li, PhD. Dissertation, Semantic-based Learning Resource Management and Exploitation, Graduate School of Chinese Academy of Science, Institute of Computing Technology, 2005.