Research on the Method of Geospatial Information Intelligent Search Based on Search Intention Model

Jingbo Liu[1, 2*], Jian Wang[2], Bingbo Gao[2]

[1] School of Automation, University of Electronic Science and Technology of China.

611731, Chengdu, P. R. China

2 Beijing Research Center for Information Technology in Agriculture. 100097,

Beijing, P. R. China

E-mail: First author: liujingbohn@163.com

Abstract: The paper, focusing on the problem that the overall correlation level of the research result is low in the current geospatial data search system, establishes users' search intention models, researches the intelligent and professional geospatial information search method, and makes use of spatial cognition theory and knowledge engineering methods to make up for the shortcomings of the lack of semantic information in the traditional keyword-based information retrieval. The geospatial information of towns and villages spatial planning field involves multitudinous theme and has complex structure and large scale. There has always been lack of appropriate method to complete the information search and processing activities. This paper takes geospatial data search in towns and villages spatial planning field for example, constructs the professional knowledge base by knowledge engineering, implements semantic reasoning and establishes users' search intentions   model by combining the characteristics of geospatial information. The ultimate goal is to understand users' search intentions as far as possible and improve the accuracy of geospatial data search.

Keywords: geospatial information search, knowledge engineering, search intention model

1   Introduction

With the advent of Internet era, the amount of available geospatial information on the Internet has dramatically grown. Compared with non-spatial information, geospatial information has more complex composition. How to improve the overall correlation level of geospatial information between search results and users' search intentions is one of the key problems to be solved.

Metadata is defined as data about data. For geospatial data, metadata can describe the content, quality, condition and other relevant background information about characteristics. The search models of existing geospatial information search systems are usually designed based on geospatial information metadata contents. These search systems match the related keywords users enter with spatial metadata database and return the results to users. To a certain degree, keywords can express users' search intentions, but this method has some shortcomings. The keywords submitted to computers are just physical sign in this method. It seriously dissevers the semantic relationship among the words users type and doesn't contain any semantic information in search process, leading to satisfactory search results [1]. Since this query mechanism is lack of intelligence and knowledge inference, it is more and more difficult to meet the growing needs.

In the field of the towns and villages spatial planning field, planners need to collect and analyze a variety of geospatial information. The geospatial information involves multitudinous theme and has complex structure and large scale. However, there has always been lack of appropriate method to complete the information search and processing activities. After analyzing the shortcomings of the traditional geospatial information search systems, the paper puts forward a semantic-based intelligent search method.

2   Intention Model

2.1   The Composition of Intention Model

In the theoretical model of intelligent agents, belief, desire and intention are regarded as the basic psychological elements of intelligent agents. It is generally believed that faith describes subject's basic view on the environment and the subject uses it to express any future possible situation. Desire is obtained directly from belief, including the subject's judgement on future situation. Intention is the goal or sub-goal chosen to achieve. In other words, the current goal or sub-goal is the subject's intention. Intention model is the formal expression of intention, that is to say how to reflect the target models users choose [2].

Geospatial data describes three basic characteristics of various kinds of phenomenon in the real world: spatial characteristic, time characteristic and thematic attribute characteristic. In this study, considering the characteristics of geospatial information, the search intention model consists of four sub-models: topic intention sub-model, measure intention sub-model, space intention sub-model and time intention sub-model.

Each sub-model has an input interface used to express users' different search intentions. Especially, in the topic intention sub-model and measure intention sub-model, the search intention users type will be submitted to inference engine. Then inference engine will reason with the knowledge base built in a knowledge engineering method. The knowledge base consists of measure ontology, planning domain ontology and inference rules.

2.2   Ontology Overview

An ontology is a formal, explicit specification of a shared conceptualization [3]. 'Formal' refers to the fact that the ontology should be machine readable, which excludes natural language. 'Explicit' means that the type of concepts used, and the constraints on their use are explicitly defined. 'Shared' reflects the notion that an ontology captures consensual knowledge, that is, it is not private to some individual, but accepted by a group. Ontology is suitable for knowledge representation, knowledge sharing and interoperability [4].

2.3   Introducing the Measure Ontology into the Measure Sub-module

This paper introduces measure ontology and inference rules to express users' measure search intentions. There is no concept of measure inference in traditional geospatial information search system. Users need to use measure knowledge, such as scale and precision, to determine whether the current data

could meet their needs. Because of the multiple form of measure, especially the complicated dependent and interactive relationship among the different measure variables, it is difficult for the majority of non-geographic experts to judge the appropriateness of the measure.

Adding measure ontology to geospatial information search system not only simplifies users' maneuverability, but also improves the accuracy and efficiency. As a new constraint variable, measure is effective in narrowing search space and enhancing the search efficiency. The core issue to be solved during introducing measure inference into search process is the inconsistency between the measure variables users type and the measure variables to be searched. Consequently, measure sub-module introduces the measure ontology and inference rules that inference engine can 'understand' and realizes intelligent search.

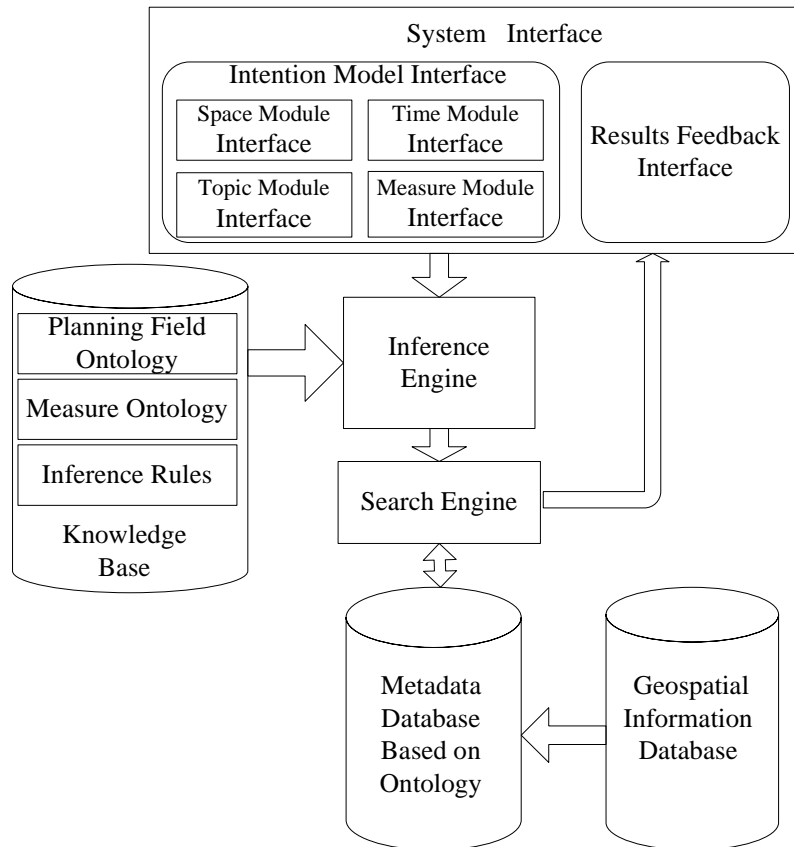### 2.4 Introducing the Planning Domain Ontology into the Topic Sub-module

The shortage of study on relevant information at the conceptual level, including the lack of information standards and domain ontology, leads to the unsatisfactory results to some degree. Using domain ontology to create semantic tagging for metadata and support semantic-based information search is a method to be mature. The development and application of the planning domain ontology will be used as the basis of the information intelligent process. This study seeks to establish like intelligent information search and process system for space planners. We firstly introduce domain ontology to assist users to express the topic intentions, and then create semantic tagging for metadata by the ontology.

3 The System Design

3.1 Overall Design

The system consists of system interface, knowledge base, inference module, search module, metadata database and spatial information database.

Figure1 shows the system framework.

Figure. 1. system framework

The basic process of geospatial information search based on the intention model is as follows. Firstly, with the help of domain experts, we construct the planning domain ontology and the measure ontology. Secondly, we create semantic tagging for the metadata of the metadata database with the two ontologies. When users type search intention through the search interface, the system will reason out relevant data with the help of the ontology and inference rules and return the results to users.

3.2   System Interface

User interface module is composed of space module interface, time module interface, topic module interface, measure module interface and results feedback interface. The topic interface and measure interface are the core interfaces of the search system. With the two interfaces, the system can fully understand the users' search intention to improve its accuracy in an intelligent and professional geospatial information search method. In the space interface, users can specify the spatial location of space objects and the spatial relationship among adjacent space objects. In the time interface, users can specify the time span of geospatial information.

Information search is the process of interacting with search system dynamically. Users often could not achieve their goals at a time while searching information. They need to option result feedback and interact with search system several times. Consequently, in this paper, the search system gets the related information by the inference engine and search engine, returns the results to users and allows users to select a second time or more.

3.3 Knowledge Base

Knowledge base is the collection of facts, rules and concepts. In terms of knowledge storage, the institution to store and manage the knowledge is called knowledge base [5]. Knowledge base system involves a key issue: knowledge representation. Knowledge representation is the key problem in knowledge base system. It should represent knowledge in the way that computers can "understand". In this paper, we realize knowledge representation with ontologies described by OWL [6]. OWL is a language proposed by the W3C consortium to be used in the "semantic-web" environment for the representation of ontologies. This language is based in the previous DAML+OIL language and it is defined using RDF. The inference rules in the knowledge base are written by SWRL [7]. The characteristic of SWRL is that it can be regarded as a part of ontology, and be supported by a lot of reasoning tools.

3.4 Inference Module

Inference engine is the core part of the inference module. It is used to drive answer from the knowledge base for the ultimate purpose of formulating new conclusions. Knowledge reasoning is the process that inference engine reasons by the existing knowledge and returns the results to users in a friendly way.

3.5 Geospatial Information Database and Metadata Database

Geospatial information database is used to store geospatial information. Geospatial metadata is a summary document providing content, quality, type, creation, and spatial information about a data set. From a data management perspective, metadata is important to maintain an organization's investment in spatial data. Metadata database is regarded as the repository into which all geospatial information metadata documents are stored. In this paper, we create semantic tagging for geospatial information metadata with the above ontologies and inference rules.

4 The Realization of the System

4.1 The Realization of the Knowledge Base

The knowledge base includes the planning domain ontology, the measure ontology and the inference rules. Establishing domain ontology is the process of abstracting and refining the objects in the real world and mapping them to the computer world [8]. When establishing planning domain ontology, we should consider planning requirements, go by background knowledge, research the information space planners need and reference the planning field standards and related metadata specifications. In order to realize sharing, the concept must comply with certain professional standards, or it will lose the significance of establishing ontology. The planning domain ontology is used to help users to express their topic intentions. Intelligent search is mainly reflected in the reasoning process with the help of the inference rules which is written by SWRL. We lay down two kinds of inference rules so that the system can achieve two kinds of reasoning. Firstly, when planners search information in any planning step, they can get various information suggestions of the step. Secondly, planners can get the upper and lower categories of the domain ontology when they select each information category.

The measure ontology in this system is composed of the following five categories: scale, spatial resolution, dimension, data accuracy and geometric dimensionality. The scale instances of the measure ontology are composed of common scales, ranging from 1:500 to 1:500 million. The spatial resolution instances consist of common resolutions. Considering the realization of reasoning, the dimension instances only include the length dimensions and the accuracy instances only include the length precisions. The various kinds of measures are not isolated, with some transformational relations. In the search system, we lays down a series of inference rules according to practical application so that users can utilize one or more instances to deduce other corresponding examples by means of inference engine. When users   type their search intentions, the system will form a measure variable constraint set according to the users' queries and return relevant data from database with the constraint set.

4.2   The Realization of the Inference Module

Jena inference engine is used as the inference engine in the search system. Jena is a Java framework for building Semantic Web applications. It provides a programmatic environment for RDF, RDFS and OWL, SPARQL and includes a rule-based inference engine. Jena is open source and grown out of work with the HP Labs Semantic Web Programme.

In the system, inference engine queries ontologies by SPRAQL (simple Protocol and RDF Query Language). SPRAQL is a query language defined for RDF data model [9]. SPRAQL model contains a collection of triples, called the basic graph pattern. SPARQL can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports extensible value testing and constraining queries by source RDF graph. The results of SPARQL queries can be results sets or RDF graphs.

4.3   The Realization of the Metadata Database

Metadata database is a database of data about data. In this study, the spatial information metadata standard is ISO19139, which is proposed by ISOTC211 group .It defines more than 300 metadata elements, being mainly used for geographic information and spatial data. The elements are divided into 14 packages, including identification information, data quality information, description information, space and time reference system information, data content information, and so on.
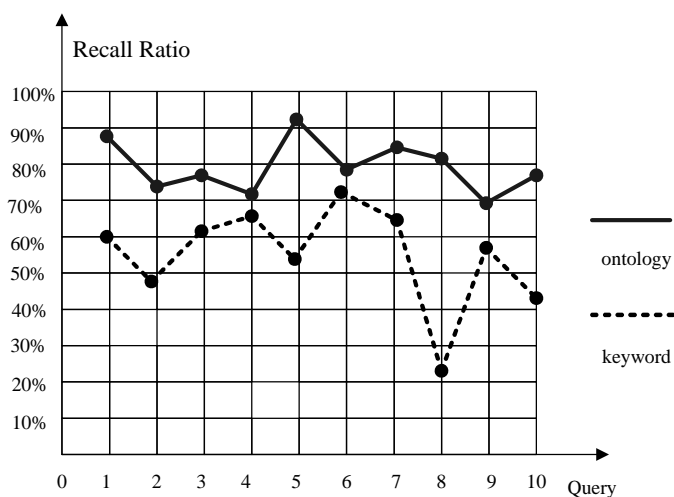
Identification information of the ISO19139 contains the topic element used to describe the phenomenon of space or space target attributes. The planning domain ontology is used to create semantic tagging for the topic element of the metadata so that the search system could support semantic reasoning [10]. ISO19139 metadata standard only contains scale and spatial resolution. It is insufficient to support the above five measure types of measure ontology. Consequently, this study expands the measure of the standard, including scale, spatial resolution, dimension, data accuracy and geometric dimensionality. Finally, we create semantic tagging for geospatial information metadata with the above mentioned measure ontology.

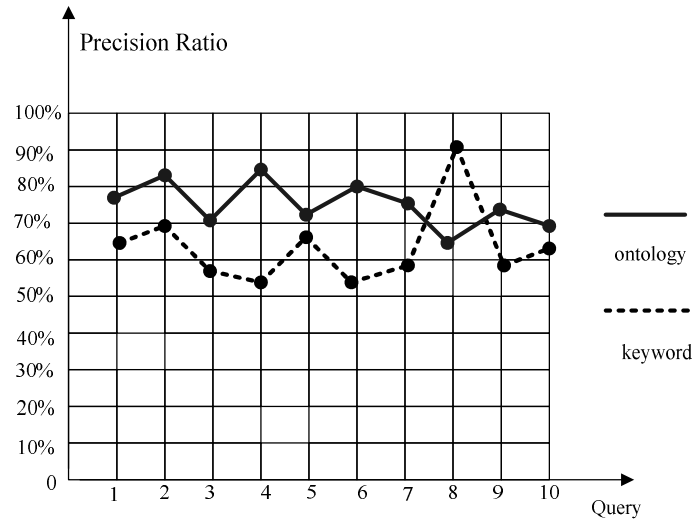# 5   The System Evaluation Result

In this test, we choose common measure and topic instances whose corresponding data exist in the database, and then judge the relativity of search results manually. Taking the maneuverability into account, this experiment only has two kinds of evaluation results: correlated and uncorrelated. This paper compares the keyword-based search with the ontology-based search (Table 1) and takes recall ratio and precision ratio as the level of the efficiency. Recall ratio is the number of records retrieved in a database search divided by the total number of relevant records in the database. Precision ratio is the number of relevant retrieved records divided by the total number retrieved in a database search.

Table. 1. Test Question Set.

Number	Module	Query Input
1	Measure Module	scale 1:10000
2	Measure Module	scale 1:1000000
3	Measure Module	precision 10.0m
4	Measure Module	dimension m
5	Measure Module	resolution 50m
6	Topic Module	population
7	Topic Module	industry
8	Topic Module	climate
9	Topic Module	highway
10	Topic Module	water resources

Figure. 2.   Recall ratio result

Figure.3. Precision ratio result

As it can be seen from Figure 2, the ontology-based semantic search is higher than the search based on keywords at the recall ratio. There is no concept of measure inference in traditional geospatial information search system, leading to a low recall ratio. But, in this search system, when users input measure needs, the search system will introduces the measure ontology and inference rules to infer user's measure search intentions and mine more implicit concept. For example, in this test, when users input 1:100000 in scale, the system will return the other four corresponding measure items. Besides, after having introduced the planning domain ontology to assist users to express the topic intentions, the search system can infer the relevant topics based on the topic users input. Therefore, the recall ration is higher than traditional method.

Figure 3 shows that the search system based on the measure ontology is higher than the search system based on keywords at the precision ratio. Adding measure ontology to geospatial information search system not only simplifies user's maneuverability, but also improves the accuracy. As a new constraint variable, measure is effective in narrowing search space. We can also conclude that the search based on the planning domain ontology is higher than the search based on keywords at the precision ratio in many instances. But, in some cases, the system appears lower precision ratio for the reason that it reasons out too much related concept. Therefore, the search system provides users a second choice to improve the precision ratio. In summary, the method is superior to the traditional method at the precision ratio.

6 Conclusions

In this paper, we firstly analyzed the deficiencies of current geospatial information search systems. Then, combining with the spatial information search of the villages and towns spatial planning field, we designed and developed a spatial information search system based on search intention model. Considering the characteristics of geospatial information, the search intention model was composed of topic sub-module, measure sub-module, space sub-module and time sub-module. In topic sub-module and measure sub-module, we established a knowledge base which includes ontologies and inference rules, and then realized the knowledge reasoning. Compared with traditional keyword-based search, this search mechanism can find implicit information and raise the recall ratio and precision ratio to

some degree. There are still some shortcomings in the search system. In future work, we will dig users' search requirements deeply and improve the inference rules.

References

1 Alon Y. Halevy.: Theory of answering queries using views. SIGMOD Record. 29(4), 40--47 (2000).

2 Rao, A., Georgeff, M.: BDI agents: From theory to practice. In: Proceedings of the 1st International Conference on Multi-Agent Systems (ICMAS-95), San Fransisco (1995)

3 Stuber R, Benjamins V R, Fensel D.: Knowledge Engineering, Principles and Methods. Data and Knowledge Engineering, 25(1-2), pp.161-197 (1998)

4 Saias, J., Quaresma, P.: A methodology to create ontology-based information retrieval systems. In: Pires, F.M., Abreu, S. (eds.) EPIA 2003. LNCS (LNAI), vol. 2902. Springer, Heidelberg (2003)

5 Yisheng Yu, Yufeng Zhang.: Knowledge Base System Study Based on Ontology. Journal of information. 22(7) (2003)

6 Owl web ontology language guide. Technical report, http://www.w3.org/TR/owl-guide/.

7 Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, MikeDean.: SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C MemberSubmission (2004).

8 Noy, N.F., D.L. McGuinness.: Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Medical Informatics: Stanford CA (2001)

9 The Sematics of SPARQL, http://www.inf.unibz.it/krdb/w3c/sparql/.

10 Cho, H., Ishoda, T.: Designing Metadata with Existing Ontologies. IEIC Tech-nical Report (Institute of Electronics, Information and Communication Engineers) (2005)