

Place Name and Address Matching Method Based on Semantic Mining

Lv Huanhuan^{1, a}, Zhang Hui^{1, b}, Wang Yuhua^{1, c}

¹ School of Software, Liaoning Technical University, Huludao 125100, China

^alvhh2010@126.com, ^bzhjordan45@126.com, ^cyuhuawang@126.com

Keywords: address matching, semantic mining, geographic ontology, inference rules.

Abstract. A place name and address matching model based on semantic mining is proposed in this paper. Firstly, the semantic features of place name and address description information is analyzed. Then, spatial object cognitive representation model geographic ontology is constructed to formalize the place name and address described by natural language. Finally, the inference rules of spatial relations are defined in order to achieve semantic matching and retrieval of referential address. Experiments show that the proposed method can accurately express and reasoning out the location information of geographical entity, and can serve the place name and address matching tasks with semantic characteristics.

1. Introduction

Geographic information public service platform[1] as a carrier of the location service brings efficiency and convenience to the urban construction, public life and economic development. Thematic information owned by the government, enterprises and other functional departments contains a large number of place name and address, and the public tend to use name and address described by natural language in the process of location query. These place name and address are all non-spatial information with semantic features[2], which need to be mapped through some techniques and methods in order to be integrated and displayed on the geographic information public service platform. Therefore, the establishment of an effective link between semantic address information and spatial information would provide a great help to the whole city information construction. So how to solve the accurate identification and precise matching of Chinese place name and address effectively is one of the important ways to promote the development of geographic information industry.

2. Related Work

The theories and methods of place name and address matching are explored and studied by researchers. In the 1970s, the U.S. Census Bureau developed Dual Independent Map Encoding (DIME)[3] and other digital geographic products achieved the real geographic distribution of code, but its structure limits the processing of geographical output results. The Topologically Integrated Geographic Encoding and Referencing system (TIGER) of the U.S. Census Bureau in 1990 introduced new address matching algorithm to improve the resolution of the geographical output. Great Britain's Address Point which is a precompiled geocoding address register and Australian National Address File (G-NAF)[4] also promoted the development of a national scale high-precision address matching functions. Many domestic researchers and scholars studied the Chinese address matching methods, which can be generally divided into rule and statistic-based, street address match and fuzzy matching. Cheng et al. proposed a rule-based segmenting and matching method for fuzzy Chinese address[5]. A city address geocoding model based on street was given by Yu et al. and matching is executed after mapping the entities such as POI and yards to the street in the model[6]. In addition, there are some other studies have also proposed different address matching methods. However, the comparison and matching of existing methods are simply from the external morphological. The semantic information of place name and address cannot be understood, and the

reasoning of relations between addresses cannot be achieved, thereby, the matching results are not enough.

3. Place Name and Address Matching Model Based on Semantic Mining

Most existing address matching methods deal with and match address strings of a standard address elements hierarchy. However, the public does not have the professional knowledge of geographic information and have different cognitive structure. As a result, the characteristics of the geographical relations and semantics of geographical object need to be given more attention.

3.1 Semantic Characteristics of Place Names and Addresses

The essence of semantic matching problems lie in different cognitions produced for the same geographic object in the process of abstraction from the real world to the conceptual world because of people's own background limitations and influences. To complete the expression and reasoning of semantic information with traditional technologies is very difficult, and the main reason is that place name and address have the following two semantic characteristics:

(1) Spatial location described by natural language has the characteristics of localization. Due to individual knowledge structure is different, the cognitive result is different. The public often use aliases and abbreviation of geographic location. Accuracy and satisfaction of matching and query results are directly affected by recognition of the public.

(2) The expressions of place name and address are also fuzzy, mainly reflected in two aspects. One is many Chinese characters denote the same referential address, which usually leads to the ambiguity of concept in the process of address matching. For example, "near" in the Chinese language can be expressed as "next to", "beside" or "side", and "opposite" as "in front" or "diagonally across" and so on. The other is the ambiguity of the geographical description of spatial information. Such as some different areas have the same place appellation, or place name is the same with azimuth name. These names are easily confused in the traditional matching way which is morphological matching, and give a barrier to place name and address matching.

According to the semantic problems, the key to solve the semantic matching is to adopt formal and clear definition expressions of geographical spatial objects and their relations, and in order to meet the cognitive needs of different background users. Geographic Ontology is used to represent the connotation of the geospatial concept and relationships in different levels and different applications, and gives semantic identification of concept. Therefore, Geographic Ontology can be used to provide semantic expression of address matching and reasoning, then realized the matching in the level of knowledge and semantic.

3.2 Geographic Ontology Matching Model Building

The OWL language and Protégé 4.3 tools are used to implement the building of Geographic Ontology. The core inference rules in semantic matching of place name and address is the rules to obtain location, including direct and indirect obtaining rules.

Direct obtaining rule of location information through Geographic Ontology: The relationship R_1 between concept C_1 and feature type is represented as $hasGeometry(?C_1, Geometry)$, and the relationship R_2 between C_1 and geographic coordinate is represented as $hasGeoref(?C_2, Georef)$. I_1 is an instance of C_1 , so it can be inferred that I_1 has the relationships of concept C_1 . Object properties are inherited directly from upper level, and numerical attributes acquire from the defined Ontology. For example, "Qian sheng department" is an instance of "mall", so its feature type can be reasoned out and its reference coordinate also can be obtained by invoking address code library in the light of numerical attributes defined in the Ontology.

Indirect obtaining rule of location information through Geographic Ontology: location information of object in the Ontology is acquired indirectly rely on topological relationship reasoning. If I_1 is the instance of concept C_1 , and it has a direct location information. Another instance of C_1 is I_2 , and there is a relationship of contain between I_2 and I_1 , so I_2 has the same location with I_1 can be inferred. For example, the coordinate of "Qian sheng department" is known, and an instance of

concept "mall" is "A", and "A" is contained in "Qian sheng department", so the coordinate of "A" can be obtained indirectly.

Matching rules of reference place names and addresses mainly include the following two:

A reference address is "the opposite of POI", the calculation process of its matching is: Firstly, the associated line of the given POI should be found. Then, if the associated line exists, the symmetric point of the given point on the line segment is the matching result. SWRL (Semantic Web Rule Language) rules are used to represent spatial calculation and relationships. The rule of above matching is shown in Figure 1, and arrows in the figure shows that the upper rules rely on the below ones.

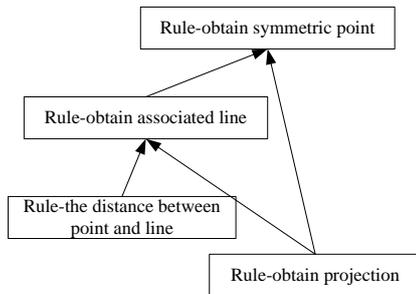


Fig. 1 Rule of "the opposite of POI"

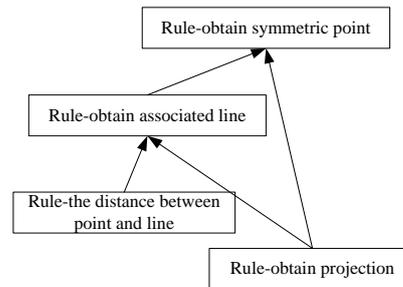


Fig. 2 Rule of a reference address

A reference address is "100 meters south of the intersection ", the calculation process of its matching is: Firstly, the associated lines of the given intersection is found. Then, get the directions of the lines and choose out the line has the consistent direction with the query direction. Finally, along the chosen line on the query direction 100 meters south from the intersection is the result point. Spatial calculation and relationships among them in SWRL rules are shown in Figure 2.

In addition to the above rules, according to different characteristics of the application, further extension can be made on the basis of spatial relationships such as topology, location and distance. Then better and richer inference rules will be defined in order to improve the accuracy of matching.

4. Ontology Matching Experimental System

4.1 Geographic Entity Matching Module

Geographical entity matching module is mainly used to obtain the geographical entity objects of corresponding matching conditions. In view of the experimental data, the following matching and query tasks are put forward:

- (1)Query of "Fuxin mining institute"
- (2)Query "buidings" on "Zhonghua Road"

In the Geographic Ontology framework constructed in this paper, query tasks above include the keywords as "Fuxin mining institute", "buildings" and "Zhonghua Road" cannot obtain matching results directly from the address code library and geodatabase. The function and significance of the Ontology are validated through the designed experimental system. The module of "geographic entities matching and query" should be selected from the experimental system main interface. This module is used to match geographic object instances under a certain condition, and then return the relevant geographic data through the instance names.

The first matching experiment: Through acronyms and aliases stored in Geographic Ontology the corresponding geographic object instances are matched. "Fuxin mining institute" is another name of "Liaoning Technical University". Since this information is stored in the knowledge base of Ontology, the needed result can get through query the keyword in the Ontology. The experimental result is shown in Figure 3.

The second matching experiment: According to the Geographic Ontology concept hierarchy, "school", "hospital" and other entities are all belong to the category of "buildings", and topological relationships between the instance "Zhonghua Road" and all the buildings are stored in the Ontology. Results satisfied the matching condition can be inferred through the relationships, and the results are shown in Figure 3.

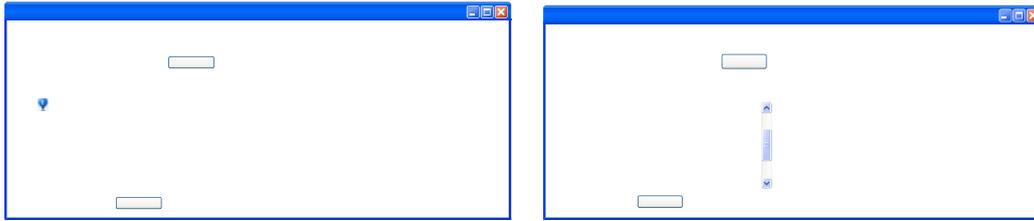


Fig. 3 Matching result of geographic entity (Experiment 1 and 2)

4.2 Geographic Entity Location Acquisition Module

This module is used to acquire specific reference coordinates of geographic entities in corresponding match conditions. The matching task is to query the location of "Labor and Social Security Building". The experimental result is shown in Figure 4:

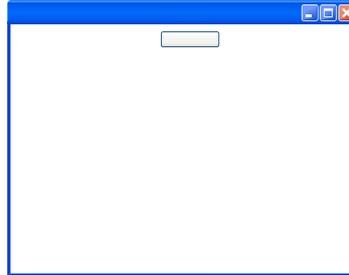


Fig. 4 Matching result of geographic entity location

5. Conclusions

Place name and address with semantic characteristics described by natural language is an important work of matching task. Geographic Ontology is introduced in this paper to express geographic objects and their mutual relationships in a formal and clearly defined way. The inference and matching of place name and address location are implemented by defining spatial calculation rules. The experiment results get from the system show that the proposed method is correct and effective, and is suitable for matching the place names and addresses with localization and fuzzy characteristics.

Acknowledgments

This work was financially supported by National Natural Science Foundation of China (61540056).

References

- [1] ZHANG Xijun, ZHANG Zhiwen and SHI Yong, "A platform for geoinformation common services data update technology-taking Shenyang digital geospatial framework as an example," *Engineering of Surveying and Mapping*, vol. 24(11), 2015, pp. 76–80.
- [2] TAN Yongmei, WANG Zhihao, "Recognition of Textual Entailment Based on Multi-Feature," *Journal of Beijing University of Posts and Telecommunications*, vol. 38, No. 6, 2015, pp. 98–103.
- [3] Janis Siksnans, *A geocoding information system for Greenland*, Technical University of Denmark, 2012.
- [4] D.W.Goldberg, J.P.Wilson and Knoblock, *From Text to Geographic Coordinates: The Current State of Geocoding*, *Journal of the Urban and Regional Information Systems Association*, vol.19, No.1, 2012, pp. 33-46.
- [5] CHENG Changxiu, YU Bin, "A Rule-Based Segmenting and Matching Method for Fuzzy Chinese Addresses," *Geography and Geo-Information Science*, vol.27, No. 3, 2011, pp. 26–29.
- [6] Yu Huanju, Li Yunlin and CAI Yongning, "A Study of Block Segmenting Method of House Number," *Bulletin of Surveying and Mapping*, vol.4, 2013, pp. 37-41.