Place Name Ambiguities in Urban Planning Domain Ontology

Yi-Pei Liao[1] and Feng-Tyan Lin[1,2]
[1]Department of Urban Planning, National Cheng Kung University, Tainan, Taiwan
[2]College of International and Cross-Strait Education, Asia University, Taichung, Taiwan

Keywords: Urban Planning Domain Ontology, Spatial Ontology, Place Name Ambiguity.

Abstract: Integrating spatial information is a crucial step in construction of Urban Planning Domain Ontology (UPDO), and taking spatial information from the web as the input of self-learning method are commonly used in constructing ontology. In this case, the use of place names could be important indicators of understanding the spatial information on the web. However, place names expressed in natural language bring diverse ambiguities, which would bring great challenges to several research fields such as Geographic Information System (GIS) and Geographic Information Retrieval (GIR). GIR has more contribution on place name ambiguities than GIS. Nevertheless, from the perspective of the urban planning domain, it is still lacking in application. This paper is a position paper that aims to bring out an argument of place name ambiguities in UPDO, and introduce two kinds of ambiguity frequently appearing in the urban planning domain. The paper also proposes a hierarchical structure of spatial ontology that allows constructors to deal with ambiguities. We believe in that the ambiguity issue is critical for urban planning, and the argument is worth discussing to all relevant domains.

1 INTRODUCTION

In construction of Urban Planning Domain Ontology (UPDO), the integration of geographic information undoubtedly is an important task. In practice, UPDO needs spatial information rather than geographic information as spatial information involves connections of locations, people, and activities (Lin et al., 2013). In view of that, place names could play an important role in integrating spatial information especially when ontology constructors need to retrieve spatial information from the web. However, more and more place names have been created rapidly and informally through various internet activities, for instance, tagging function on community websites. Matching a place name to a real space therefore turns more ambiguous than before as a place name is more likely to become a vague concept instead of a geographic information. Moreover, some of conceptual place names are widely used in formal documents, such as press releases and formal governmental reports. This brings out a new challenge for UPDO to deal with the ambiguity.

In fact, place name ambiguities are not a new issue. The research field of Geographic Information System (GIS) has studied on this question since the 1990s, and so did the field of Geographic Information Retrieval (GIR), a technique combining the information retrieval and spatial ontology. GIR has contributed to derive the geographic information from web documents automatically, as a result, recognizing the place names from natural language is one of the great challenges in GIR. Nevertheless, the existing disambiguating technique in urban planning domain remains insufficient. More efforts are needed to be put for solving place names ambiguities while constructing UPDO which is grounded on robust spatial ontology.

This paper is a position paper bringing out an argument over place name ambiguities. It presents two kinds of ambiguity that are commonly seen in urban planning domain, yet have not been taken into consideration in GIR. In addition, it proposes a new hierarchical structure of spatial ontology which is able to deal with the ambiguity. The paper is organized as follows. Section 2 will review works with regards to place ambiguities in GIS and GIR. Section 3 introduces two kinds of place name ambiguity that have not been dealt adequately in terms of urban planning practices. Section 4 brings out a new structure of spatial ontology which considers the place name ambiguity discussed in

Liao, Y. and Lin, F..
Place Name Ambiguities in Urban Planning Domain Ontology.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 2: KEOD*, pages 429-434
ISBN: 978-989-758-158-8

Section 3, and proposes a brief outline for proceeding to construct the spatial ontology. Finally, conclusion is given in Section 5.

2 BACKGROUND

In the following are reviews of works regarding place name ambiguities. The first part will firstly focus on the field of GIS, and then shift the focus to the spatial ontology and GIR.

2.1 Place Ambiguities in GIS

Name of place is a basic attribute of location information in common spatial database, accompanied with other basic attributes like latitude, longitude, altitude, coordinates, area, etc. Most of spatial databases advocate an entity-oriented view of space, which means that the space data which could be in the form of a point, a line, or an region is an exact object. Some difficulties are discovered when it comes to defining an area with indeterminate boundaries (Burrough and Frank, 1996; Wang and Hall, 1996). One of difficulties is a *vague region* or *fuzzy spatial data types* (Erwig and Schneider, 1997; Schneider, 2008). Fortunately, many researches and tools have advanced in dealing with the problem of indeterminate boundaries and vague regions.

Fuzzy region problem has been handled in GIS, however, the spatial data representing through the open-data and internet platform is still based on the traditional view of space, for example, Google Map and Open Street Map (OSM). Furthermore, many map tools commonly used by users and netizens display the place data through a standard format, such as KML or GML, yet the simplicity will bring difficulties in recognizing the place from expression.

2.2 Place Ambiguity in Spatial Ontology and GIR

Despite of the advancement in GIS, spatial ontology development and knowledge management in dealing the ambiguity problem remains in a primary state. Numerous works have put efforts on constructing ontology, but few of them have focused on the time and space dimension of thematic ontology. Peuquet (2001) worked on an ontology framework which could derive effectively the what/when/where information with robust space-time data structure. Also, he developed the query language, the operation, and the users interface. Perry et al. (2006) drew an outline of basic classes and relationships for a spatial upper-ontology, which brought spatial dimension into other ontology and allowed spatial query operation.

The technique of Geographic Information Retrieval (GIR) has focused on the spatial relations between any kinds of knowledge, where information is described with geographic metadata. GIR is user-oriented applications including spatial query, search and display functions. For example, the Spatially-Aware Information Retrieval on the Internet (SPIRIT) (Jones et al., 2002; Jones et al., 2004) is a search engine for geographic information. SPIRIT advocates that most of the web resources refer to geographic space, which means that the event was recorded once when it reveals or happens in a certain place. SPIRIT regards each entity as a geography entity with geographic information. In the part of query method, SPIRIT intelligently understands the users' searching language and tells any possible event that has relation with geographic information. Mata and Claramunt (2011) rested on the contribution of SPIRIT and gave an approach for retrieval of geographic entities according to its spatial, temporal, and thematic information. The approach extracted diverse dimension of information from the gazetteer, which has its own XML format of geographical entity (eg. Wikipedia).

Several challenges in GIR demand further significant researches (Jones and Purves, 2008), including (1) detecting geographic reference in the form of place names and spatial natural language, (2) disambiguating the place names, (3) indexing documents respecting to their geographic context, (4) ranking relevant documents with respect to geography as well as theme, (5) developing effective user interfaces, and (6) developing methods to evaluate the success of GIR. The research introducing in this paper focuses on the first two challenges, the place names detecting and place names disambiguating. This research also faces another difficult but critical challenge of the language characteristics in Chinese.

In summary, GIS and GIR are both very fundamental tools in the domain of urban planning. In GIS, there is advancement in representing places even though data might be either specific or vague. The problem of vague region is much like a semantic problem of space, and ontology is considered as a major method to deal with the semantic problem. However, it is still in a primary stage in terms of spatial ontology and GIR.

3 PLACE NAME AMBIGUITY

The nature language problem in different languages is important to GIR, and the place name ambiguities is one of top priorities needed to be dealt with, and a lot of researches have put efforts on it. Nevertheless, only a few of place names ambiguity researches have dealt with the problem that would happen in using Chinese language. Chinese characters, different from English word, is defined separately by intrinsic meaning rather than by visible space, resulting in that a character cannot be recognized by a normal pattern. Furthermore, in Chinese the meaning of a character could become very different when combining with other characters. Thus, defining a character and a term is a critical task in Chinese as its meanings are strongly dependant on the grammar analysis system. The follows are going to introduce two kinds of Chinese place name ambiguity that often appear in urban planning domain.

3.1 The Place Name as a Concept of Spatial Distribution

A planner may need to query for location information of specific spatial events, such as “flooding area” and “potential flooding area”. In this case, there are at least two ambiguity problems that make “flooding area” unable to be identified: (1) how can the character “flooding” be recognized as a part of a place name? It is more likely to be identified as an adjective. How to distinguish an adjective for a place name from an adjective in normal would then become a problem. (2) The term “flooding area” refers to a concept of geographic distribution instead of a certain location. However, according to the assumptions of normal GIR, there could be only one focusing place in a single place name entity, hence “flooding area” is unrecognized because it is a description of a spatial event in name yet without a continuous distribution in space.

3.2 The Place Name as a Concept of Social Phenomenon in Space

Another place name ambiguity is caused by conceptualized social phenomena. This kind of place names could be a proper noun individually that is easy to be recognized by pattern. Nevertheless, it also has a very fuzzy boundary or even has no boundaries on a map. The reason is that these place names are created for describing particular social development phenomena such as poverty gap and real estate. In other words, the phrases have been named before being indicated in space. In addition, these place names could be unprofessionally defined as they may be created by the public, especially the netizens.

One of typical examples for this type of ambiguity is a compound term called “Tyan-Long Nation”, which is firstly created by some netizens and is now frequently used by news media in Taiwan. “Tyan-Long Nation” has an ironic meaning initially that indicates a place where its residents are self-centered and ignorant about anything happened in other places. Yet nowadays the meaning has evolved into a phrase describing a place with high commodity price and extreme high housing price. Moreover, verbally “Tyan-Long Nation” refers to the Taipei City, the capital of Taiwan, however, they are not exactly matched geographically.

In planning domain ontology, it is necessary to understand this type of place names. Remind that it is probably hard to find “Tyan-Long Nation” on the Google Map because it is a phrase for a specific concept rather than a location. Furthermore, it is a name created by the public rather than by the official. Therefore, in order to reveal the fuzzy boundary of “Tyan-Long Nation”, we collect all locations with a name “Tyan-Long Nation” from the Facebook places, in which locations are allowed to be created by all users, and check how much Facebook users have checked in at each location. The more the location has been checked, the higher score as well as the possibility is assigned to the location. Figure 1 shows the result of this survey. The boldest black line is the boundary of Taipei City, and the grey circles are all locations that are named with “Tyan-Long Nation” each circle has the radius about 1.2 km, and with shade degree base on the number of

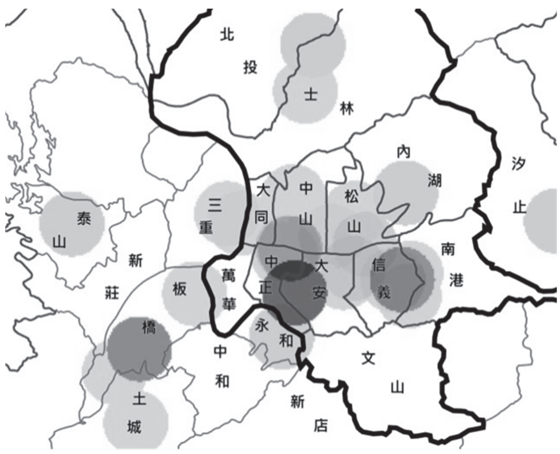

Figure 1: A map displays the “Tyan-Long Nation” according to the mass Facebook users.

checked users. Figure 1 clearly reveals that "Tyan-Long Nation" is concerned with some places in Taipei City but not all. In this survey, the place name was created basing on the social phenomenon with a vague region, and we display the region on a map according to the diverse understandings of mass Facebook users.

4 SPATIAL ONTOLOGY FOR PLANNING DOMAIN

In order to develop a spatial ontology for urban planning GIR and UPDO, we need a new structure of spatial ontology that is able to handle the ambiguity problems identified in Section 3. In the following part will firstly introduce the source of place names, then a new structure for spatial ontology, and finally a framework of semi-automatic construction of this spatial ontology.

4.1 Place Names from Facebook

Facebook is the major sources of place name collection in this research. The contents are generated from the users' perspectives, thus how to take these informal contents into scientific and theoretical research becomes a critical point, which could be a very fundamental question. In the case of collecting place names from Facebook, it's obvious that not every check-in name could be regarded as a place name, and in contrary, it's also expected that some informal check-in name should be identified as a place name. The critical and fundamental question is "what is a place name?"

Cresswell (1996) aimed that the concept of place should be scrutinized to both geography and human everyday life. A place name is created only when the place have some relations with some human activities. In the domain of urban planning, the rethinking of the relation between space, human, and symbols (names) is quite associated to the concept of "City image," which was brought out by Lynch (1960). City image argues that a city's space is not defined by its structure design but by the feeling of the people living in it. City image is a mental map of a person, and it might be very different to each person even though they are experiencing in the same city. Based on the concepts above, it's interesting to look through the Facebook place names and analysis the reason of why the name is created, is it related to "what people feel about the place?" or "what people do at the place?"

Figure 2 shows the procedure of Facebook place names Extraction. There are four parts between the extraction from web and the storing to database: (1) API is a tool to collect data from Facebook by using the Graph API. So far there are about 10,000 place names in Taiwan have already collected from Facebook. The metadata has 10 items: name, category, street, city, state, country, zip, latitude, longitude, and check-in. The check-in is the count number of users who checked in with that place name. (2) Potential place name extraction is working on analysing the name and deciding whether it is possible place name or not. There are several kinds of situation that make the name none potential, for example, there's a name called "on a moving train" or "car racing." None potential place names are regarded as unidentifiable names which will leave to (3) place name re-identification process. In this process, several algorithms are developed for different unidentified situations. (4) Place name formalization is the last step before storage. Formalization will deal the structural ambiguity problems, such as shortened names and alternated names. It's based on the previous work by Deng et al. (2012), which has developed an algorithm to extract Chinese place name by using natural language processing (NLP) method.

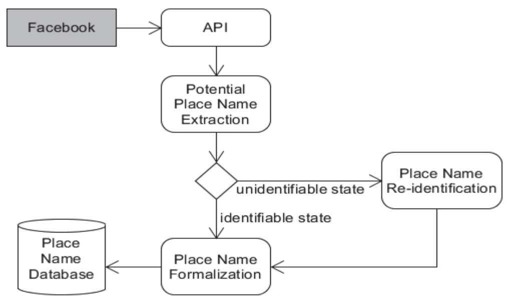

Figure 2: The procedure of Facebook place name extraction.

4.2 A Structure of Spatial Ontology

An ontology-based model is utilized in the new defined spatial ontology. Figure 3 shows the UML class diagram of the hierarchical spatial ontology, which allows the situations mentioned in Section 3. The place is divided into three levels. The *place* class at the first level refers to a general concept. Following the *place* class, *normal place* and *distribution area*, stand at the second level. The *normal place* bases on the prototype and algorithm in previous research of GIR. The *distribution area* is separated in two sub-categories, namely, *spatial event* and *social event*, which are both learned from

UPDO. The *distribution area* could have *distribution rate* relations with several *normal places*, where each *distribution rate* describes the frequency of the *distribution area* has distributed in that *normal place*. The *distribution rate* relation has a value *Di*, $0 \leqq i \leqq n$, where *n* is the number of total *normal place* that have *distribution rate* relations with that *distribution area*.

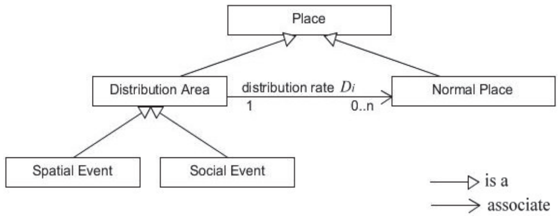

Figure 3: The class diagram of spatial ontology.

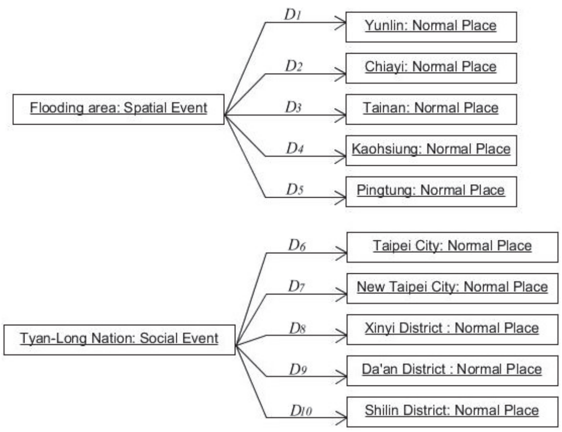

Figure 4: Instance diagrams of two kinds of *distribution area,* where "flooding area" is a *spatial event* and "Tyan-Long Nation" is a *social event.*

Figure 4 are instance diagrams of two examples mentioned in Section 3.1 and 3.2. The "flooding area" is a *spatial event* and "Tyan-Long Nation" is a *social event*, yet both of them are belong to *distribution area.* "Flooding area" distributed in "Yunlin", "Chiayi", "Tainan", "Kaohsiung", and "Pingtung", while "Tyan-Long Nation" is dispersed to "Taipei City", "New Taipei City", "Xinyi District", "Da'an District", and "Shilin District". The distribution relations are recorded respectively as symbols D_1 *to* D_{10}. These *distribution rate* values can be calculated by spatial analysis methods such as proportion of area, or by textual analysis methods such as co-occurrence rate, or by integrating the former two methods.

4.3 Semi-automatic Spatial Ontology Constructing

Based on the class hierarchy of place in Figure 3, we construct a spatial ontology with three parts, the normal place database, the distribution area database, and the distribution relations database. Figure 5 is an outline of semi-automatic construction procedure of spatial ontology, in which the rectangle with bold dotted line describes the spatial ontology.

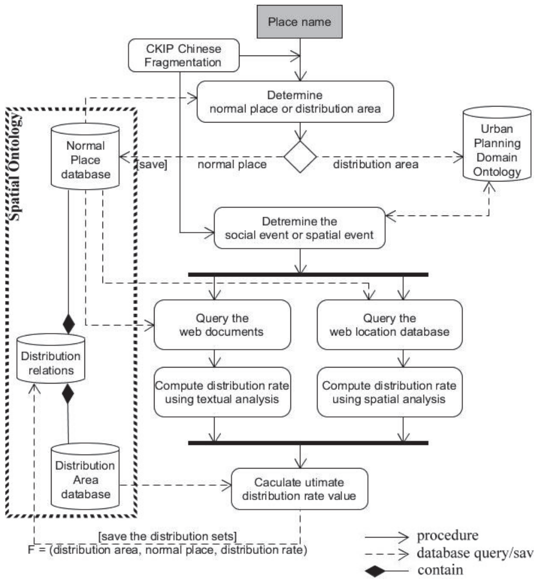

Figure 5: Semi-automatic spatial ontology construction framework in UPDO.

Figure 5 shows the procedure of semi-automatically constructing spatial ontology, and there are three points worth being mentioning: (1) A Chinese grammar analysis tool called CKIP, developed by Academia Sinica in Taiwan, would fragment terms into sub-terms according to their part of speech (Ma and Chen, 2003). CKIP involves in the process of determining the category of place from the first level to the second level. (2) The existing UPDO is able to assist in identifying any relevant spatial or social event that are embedded in the place name, and thus would help to decide the category of distribution area from the second level to the third level of spatial ontology. (3) The distribution rate is calculated according to both spatial analysis and textual analysis methods.

5 CONCLUSION

This paper is a position paper that brings out an argument of place name ambiguities. Although some ambiguities have been taken into account in GIS and GIR researches, the domain of urban planning, which especially needs to integrate all knowledge existing in a particular space, still lacks for the consideration. Two types of ambiguities in Section 3 are the evidences indicating a wide gap between the place name and the physical space. We also propose a new defined structure of spatial ontology that will be utilized in UPDO in further researches. The spatial ontology presented in Section 4 is a fundamental framework for urban planning GIR and UPDO. We believe that the contribution of this research in further can serve in several tasks such as Decision Support System (DSS), knowledge understanding, and the automatic learning of relevant domain ontology.

REFERENCES

Burrough, P. A., and Frank, A.(1996). *Geographic objects with indeterminate boundaries,* vol. 2. CRC Press.

Cresswell, T. (1996). *In place-out of place: geography, ideology, and transgression*. U of Minnesota Press.

Deng, D. P., Chuang, T. R., Shao, K. T., Mai, G. S., Lin, T. E., Lemmens, R., ... and Kraak, M. J. (2012). Using social media for collaborative species identification and occurrence: issues, methods, and tools. In Proceedings of the *1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information* (pp. 22-29). ACM.

Erwig, M., and Schneider, M. (1997, January). Vague regions. In *Advances in Spatial Databases* (pp. 298-320). Springer Berlin Heidelberg.

Jones, C. B., Abdelmoty, A. I., Finch, D., Fu, G., and Vaid, S. (2004). The SPIRIT spatial search engine: Architecture, ontologies and spatial indexing. In *Geographic Information Science* (pp. 125-139). Springer Berlin Heidelberg.

Jones, C. B., and Purves, R. S. (2008). Geographic information retrieval. International *Journal of Geographic information Science*, 22(3), pp.219-228.

Jones, C. B., Purves, R., Ruas, A., Sanderson, M., Sester, M., Van Kreveld, M., and Weibel, R. (2002). Spatial information retrieval and geographical ontologies an overview of the SPIRIT project. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 387-388). ACM.

Lin, F. T., Liao, Y.P., and Lin, C.A. (2013) Using Ontology in Planning Knowledge Management System, the 5th Joint AESOP-ACSP Congress, Dublin, Ireland.

Lynch, K. (1960). *The image of the city*. MIT press.

Ma, W. Y., and Chen, K. J. (2003). Introduction to CKIP Chinese word segmentation system for the first international Chinese Word Segmentation Bakeoff. In *Proceedings of the second SIGHAN workshop on Chinese language processing*, volume 17 (pp.168-171). Association for Computational Linguistics.

Mata, F., and Claramunt, C. (2011). GeoST: geographic, thematic and temporal information retrieval from heterogeneous web data sources. In *Web and Wireless Geographic information Systems* (pp. 5-20). Springer Berlin Heidelberg.

Perry, M., Hakimpour, F., and Sheth, A., 2006. Analyzing theme, space, and time: an ontology-based approach. In Proceedings of the *14th annual ACM international symposium on Advances in geographic information systems* (pp. 147-154). ACM.

Peuquet, D. J. (2001). Making space for time: Issues in space-time data representation. In *GeoInformatica*, 5(1), pp.11-32.

Schneider, M. (2008). Fuzzy Spatial Data Types for Spatial Uncertainty Management in Databases. *Handbook of research on fuzzy information processing in databases*, 2, pp.490-515.

Wang, F., and Hall, G. B. (1996). Fuzzy representation of geographical boundaries in GIS. *International Journal of Geographic information Systems*,10(5), pp. 573-590.