Ambiguous Address Locating in Emergency Response

Kunwang Tao, Liang Wang,   Xianlin Qian

Chinese Academy of Surveying and Mapping

e-mail:{taokw, wangliang, qianxl}@casm.ac.cn

Abstract

In Emergency response system, it is necessary to collect thematic data from various sectors, such as data about emergency services from hospitals, schools, fire department and warehouse, especially the exact location of these facilities. Usually we cannot get geographical coordinates directly from collected data. Place names and addresses in natural language with ambiguity are more common in many scenarios. In order to integrate these thematic data into geospatial maps to get a comprehensive view of the situation of the emergency, it is desirable to transform the text format addresses into geographic points or regions. This paper explains the implementation of the geocoding process, in which ambiguous addresses are inputs and possible geographical coordinates or regions are outputs.
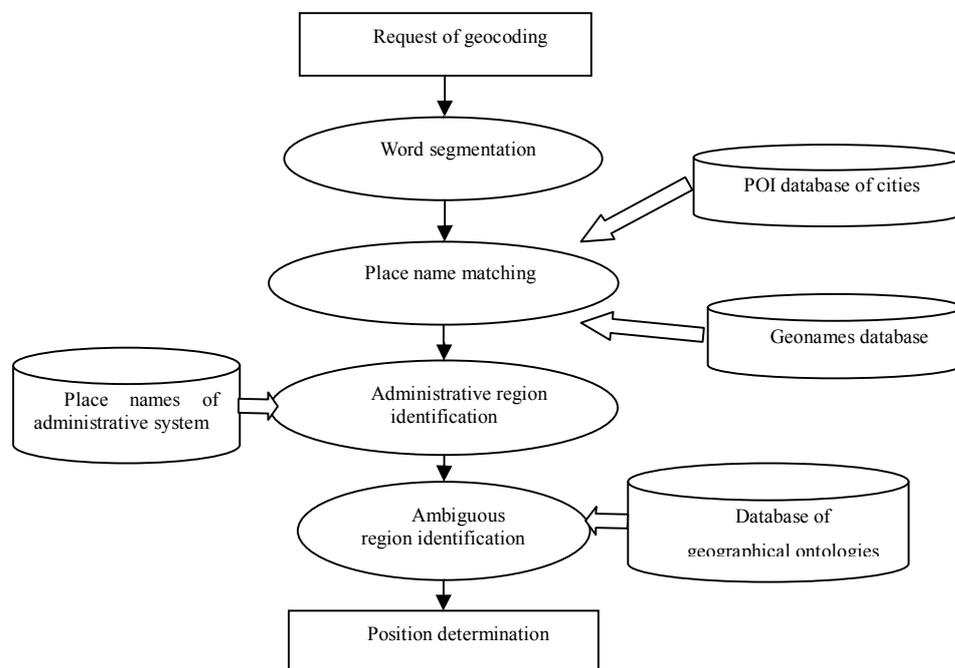
Figure 1. Flowchart of ambiguous geocoding

An ambiguous address matching system has been implemented to provide the core function of geocoding. The overall process of the system is depicted in Figure-1. The key components of the ambiguous address matching system are four databases which are used to facilitate the place name matching. The four databases are place names of administrative system (the 3 tiers hierarchical system of china' administrative division consist of provinces, metropolitans and counties), POI database of cities, Geonames database and database of geographical ontologies. Some other auxiliary data are also used in this process such as street names and gazetteers.

The process of the address locating decomposes into five steps which are word segmentation, place name matching, administrative region identification, ambiguous region identification and position determination.

Address expressed in Chinese language has the characteristics that the significance of the elements in the address is descending. Using word segmentation algorithm, the elements of the address, which include name of province, metropolitan, county, street and route number, can be extracted. There are many duplicated place name in Chinese address. The duplications are main obstacles for getting an accurate and precise geographic position for an address. To resolve this problem, following procedures have been taken:

1. According to the hierarchical structure of the administrative system, a 5-levels tree structure is built with each node represent a specific administrative region. The parent-children relation-ship in the tree corresponds to the belong-to relationship between administrative regions.

2. Traverse the tree of administrative regions to gradually positioning the elements of the address. For instance, an address of "28 lianhuachixi road, Haidian district, Beijing" can be separated into 3 elements using word segmentation algorithm. Determine the place name of province level administrative region from the tree with "Beijing", then find "haidian district" and "lianhuachixi road" down the tree.

3. If a top level element of an address is missing, then we need to check all the possibilities to distinguish the one with most likelihood by traverse the tree of administrative regions with multiple top-down searches.

4. For the descriptive text in an address which is hard to positioning such as "south to road A", "middle in road B", "300 meters to northeast", A database for geological ontologies is used to reasoning the geometric relation-ship between the geographic features and address elements. For an ambiguous address, the unambiguous elements are extracted and separated. A geographical ontology is established by this part and the ambiguous elements of the address are reasoned based on the ontology.

In experiments, more than 14000 addresses are processed to test the effectiveness and efficiency of the place name matching system. The accuracy of the derived coordinates is divided into 7 levels, which are province level, metropolis level, county level, town level, street level, road number level and building level. Results show 10% have the accuracy of building level, 65% of road number level, 75% of street level, 90% of town level, 99% of county level which means less than 1% addresses are located into a wrongly county.

Reference:

Z. Song, "Chinese address match algorithm based on natural language understanding", *Journal of Remote Sensing in chinese*, 2014:4

J. Liu, B Li,. "An ontology driven geospatial event related information retrieve approach". *Journal of Surveying and Mapping in Chinese*. 2011(40)4

LiBin,ShiLihong, "Research on Geo-event Ontology Construction Oriented Task，*2011 International Conference on Network Computing and Information Security*

C. Sun, S. Zhou, L. Yang,. "Chinese geocoding based on hierarchical place name database", *Applications of computer in chinese*. 2010(7)

H. Guo, G. Song, L. Ma, S. Wang,. "Design and Implementation of geocoding system". *Computer Engineering in Chinese*. 2009(01)

L. Zhang, S. WU,. "Design and implementation of address match engine in geocoding system". *Journal of Geomatics, in Chinese*. 2008(06)