

The Research on the Describing Knowledge of GIS Path Based on Natural Language

Abstract. The describing method of GIS path based on natural language becomes a focus of researches on the application of intelligent GIS. The knowledge of description of path is the core problem and fundament of realization. According to the classified method of knowledge which classification includes the fact knowledge, the rules knowledge and the controlling knowledge, the paper analyses and concludes all kind of description knowledge during the describing process of path, and builds the describing knowledge base of path. Finally, the paper realizes the prototype describing system with path chart and Chinese by which the describing knowledge are test.

Streszczenie. Zaproponowano inteligentny system gromadzenia danych dla GIS (Geographic Information System). Bazą jest znajomość ścieżki. Artykuł analizuje wszystkie możliwe metody opisu i buduje bazę wiedzy. (Analiza bazy wiedzy dla systemu GIS opierająca się na języku naturalnym)

Keywords: natural description of GIS path; describing knowledge; knowledge base of path description

Słowa kluczowe: system GIS, baza wiedzy, język naturalny.

Introduction

Because of according to the recognizing habit of humankind, the description of GIS path based on natural language (DPNL) not only makes describing path vividly, but also supplies users accurate and efficient information of path. The querying operation of GIS path based on natural language which used simply and friendly is more suitable to ordinary users. The description of GIS path based on text is so practical and valuable for the mobile ends (such as GPS car navigation systems, smart phones) which processing ability is weak. Therefore, the description of GIS path based on natural language has been a hot research on GIS. The core problem of DPNL is the research on describing knowledge such as describing words of path, describing sentence rules of path, and so on. Hoepfner put forward the path describing way based on basic path unit and its related information [1]; Liu Yu defined the grammar of NLRP (Natural Language Representation of Path) of restricted Chinese based on syntax model of NLRP [2]; Zhang Xueying described the structure and content of Chinese describing vocabulary base of path and designed the describing unit and the compound syntactic pattern according to characteristic of Chinese [3]; Through analyzing the researches before, two aspects still need be improved and enhanced as following: one is that the description language is relatively rigid and not very well close to the natural language; the other is that the description of the path intelligibility is poor, the lack of street signs and other auxiliary surface features indicative of information. The main reason of those shortages is lack of the researches on describing knowledge of GIS path which include the words used to describing are not enough and the patterns of describing sentences are simple. According to the classified method of knowledge which classification includes the fact knowledge, the rules knowledge and the controlling knowledge, the paper analyses and concludes all kind of description knowledge during the describing process of path, and builds the describing knowledge base of path. Finally, the paper realizes the prototype describing system by which the describing knowledge is test.

The Fact Knowledge of DPNL

The fact is the value or status of property of objective things. It is the static knowledge which is shared and recognized by humankind. And so on, it is the low-level knowledge in knowledge base. The fact only describes the real value of property or state under certain condition and

does not include any variables ordinarily. It is expressed by a proposition which value is true or by a statement of state [4]. The fact knowledge involved in the process of DPNL includes describing words of path and names of landmark features.

A. Describing Words of Path

Words are the fundament unit of natural language sentence. Describing words of DPNL are used to describe the information of path. Whether the describing result of path is practical and vivid directly relates with whether the describing words are rich and overall. Based on analysis of sentences of DPNL used ordinarily, the paper divided the words into four categories which include Verbs, Orientation words, Distance words and auxiliary words according to their function in the path describing sentence.

1) Verbs: Include special words which represent several action of walking. These words are illustrated in Chinese as follows: "步行"(walk), "行走"(go), "走"(walk), "行至"(walk to), "到达"(arrive), "转"(turn), "拐"(turn), "转弯"(turn), etc. Verbs can be chosen flexibly in the same meaning words.

2) Orientation words: Used to describe orientations of the path are divided into two classes: absolute orientation words and relative orientation words. Orientation words in Chinese include "东"(east), "南"(south), "西"(west), "北"(north), "东南"(southeast), "东北"(northeast), "西南"(southwest), "西北"(northwest), "左"(left), "右"(right), "前"(forward), "后"(behind), etc.; The orientation of nodes in road network is determined by specific algorithm. It is determined by polar coordinate in the paper.

3) Distance words: Used to describe length of the road arc include the common numerals and distance units. Such as these words in Chinese "米"(m), "公里"(km), etc.

4) Auxiliary words: Are other sentence elements of the DPNL. The words could not be divided further because types of their part of speech in Chinese are so rich and they could only play a supporting role. Auxiliary words include these words in Chinese: "继续"(continue), "然后"(then), "那么"(then), "接着"(then), "紧接着"(then), "再"(then), "此时"(at the time), "这时候"(at the time), "此刻"(at this moment), "朝着"(towards), "朝"(towards), "向"(towards), "沿着"(along), "往"(to), "穿过"(across), "途经"(by way of), etc. An auxiliary word can be replaced by other same meaning words. Auxiliary words can be refined and confirmed in the describing rules base of path.

B. Names of Landmark Features

The landmark features are used to affirm right path and guide users in sentences of DPNL. Especially in describing of the long-distance path, they are even more essential. Therefore, Landmark features' names become an important part of the describing knowledge base. The landmark features normally include landmark buildings, local well-known entities (such as schools, banks, government agencies), and so on. They can be stored and used by extending an attribute of landmark features on geographic entities.

Landmark features' name base includes the following attribute columns:

NAME: The column stores the name of landmark features. Name of the original spatial entities should be used directly in order to ensure consistency between landmark features and original spatial entities.

LAYERIN: The column records the map layer which store spatial entities of landmark features. The spatial information and attribute information of landmark features can be got in corresponding map layer data.

ID: ID is an identification of landmark features in the map layer data through which the corresponding geographic entity can be accessed from the spatial data file or the spatial database.

ARCSID: ID collection of arcs indicated by landmark features is stored in the column. Constructing the correlative relationship between landmark features and arcs in road network is realized in the algorithm of building road network topology. One landmark feature can indicate one arc of road, sometimes, can indicate several arcs of road. For example, a landmark feature locating a crossing can indicate several arcs of road. In the paper, the algorithm extended constructing process of the relationship between landmark features and arcs in road network is not specially illustrated.

The Rule Knowledge of DPNL

The rule knowledge is causal relationship knowledge related actions of objects. It is dynamic knowledge^[5]. There are many describing rules which build structure of describing sentences in the process of DPNL. The describing rules are very important for building sentence of DPNL. Road nodes, road arcs and landmark features are three basic features which compose a sentence of DPNL. According to different effect in DPNL, road nodes include four sub-types which are start node, non-direction-changed node, direction-changed node, end node. The describing rules include four types rules which are describing rules of different type nodes, describing rules of adjacent arc, describing rule of landmark features and grouping rule of syntagms. The paper describes describing rules using the logical describing method of production which be illustrated by Backus-Naur Form (BNF)^[6].

A. Describing Rules of Road Nodes

1) Describing Rule of Start Node

<Describing rule of start node>::= “您从” <start node> “出发，” <describing rule 1 of adjacent arc>

The rule defines the clause pattern which you will say normally when you set out from the starting point. In the rule, “您从” and “出发，” are special Chinese clause or word which compose the whole rule clause. <start node> expresses first node of path. <Describing rule 1 of adjacent arc> will be described in section “III.B”. A Chinese instance of the rule is “您从测绘学院出发，沿陇海路向东走，”，expressed in English as “You can start off and go along Longhai Road towards East from the Institute of Surveying and Mapping，”.

2) Describing Rules of End Node

<Describing rule 1 of end node>::= { “到” | “到达” } <crossing> “时，” <describing rule of last arc>

<Describing rule 2 of end node>::= “这” { “时” | “时候” } “到了” <crossing> “，” <describing rule of last arc>

<Describing rule of last arc>::= { “继续” | “再” } { “前行” | “向前行” | “向前走” } [“大约”] <length of last arc> [“左右”] { “便” | “就” } “到了目的地” <end node> “。”

The rules define the clause pattern which you will say normally when you arrive at the terminus. In the rules, <crossing> expresses node's name named using method in section “IV.A”; <Length of last arc> should be displaced the numerical value of length of last arc; <End node> expresses the name of last node of path. A Chinese instance of the rule is “到大学路淮河路路口时，继续向前行大约 200 米便到了目的地古玩城。”，expressed in English as “when arriving at crossing of Daxue Road and Huaihe Road, you can walk on about 200 meters to arrive at the terminus, the Market of Curio.”.

3) Describing Rules of Non-direction-changed Nodes

<Describing rule 1 of non-direction-changed nodes>::={ “在途中” | “中途” } “您” { “会” | “将” } { “看见” | “看到” | “经过” } <non-direction-changed node> “，” <describing rule 2 of adjacent arc>

<Describing rule 2 of non-direction-changed nodes>::= “到达” <non-direction-changed node> “，继续” [“向”] { “前行” | “前走” } “，” <describing rule 2 of adjacent arc>

The rules define the clause pattern which you will say normally when you arrive at a middle point of path at which the direction of path is not changed. In the rules, <non-direction-changed node> will be displaced with node's name of current node of path; <describing rule 2 of adjacent arc> is expressed in section “III.B”. A Chinese instance of the rule is “在途中您会经过陇海路与兴华路交叉口，接着沿陇海路继续前行，”，expressed in English as “you can move on along Longhai Road when you arrive at crossing of Longhai Road and Xinghua Street.”.

4) Describing Rules of Direction-changed Nodes

<Describing rule 1 of direction-changed nodes>::= “在” <direction-changed node> [“向”] <polar direction> { “转” | “转弯” } “，” <describing rule 2 of adjacent arc>

<Describing rule 2 of direction-changed nodes>::= [“当”] { “到” | “到达” | “行至” } <direction-changed node> “时” [“向”] <polar direction> { “转” | “转弯” } “，” <describing rule 2 of adjacent arc>

The rules define the clause pattern which you will say normally when you arrive at a turn point. In the rules, <direction-changed node> will be displaced with node's name of current node of path; <polar direction> should be Chinese orientation words judged using the method in section “IV.B”; <describing rule 2 of adjacent arc> is expressed in section “III.B”. A Chinese instance of the rule is “当到陇海路与大学路交叉口时向右转，接着沿大学路继续前行，”，expressed in English as “when you arrive at crossing of Loanhai Road and Daxue Road, you should turn right and walk on along Daxue Road.”.

B. Describing Rules of Adjacent Arc

<Describing rule 1 of adjacent arc>::={ “上” | “沿” | “沿着” } <adjacent arc> { “向” | “朝” } <polar direction> “走，”

<Describing rule 2 of adjacent arc>::= { “然后” | “接着” | “紧接着” } { “顺着” | “沿” | “沿着” } <adjacent arc>[“继续”] { “前行” | “向前走” } [“大约” <length of adjacent arc><Chinese words of length unit>] “ , ”

The rules define the clause pattern of adjacent arc between two middle nodes of path. In the rules, <adjacent arc> expresses the name of current arc of path which is adjacent to current node of path. In the rule 1, <polar direction> should be Chinese orientation words judged using method in section “IV.B”. In the rule 2, <length of adjacent arc> should be displaced the numerical value of length of current arc; <Chinese words of length unit> is Chinese word expressing length unit. The rule 1 is often used in <describing rule of start node>; the rule 2 is often used in describing rule of middle node such as non-direction-changed node and direction-changed node. A Chinese instance of the rule is “继续沿着陇海路向前走大约 200 米, ”, expressed in English as “go on along Longhai Road for about 200 meters.”.

C. Describing Rule of Landmark Features

<Describing rule of landmark features>::= “途中您” { “可能” | “将” | “会” | “将会” } { “看到” | “看见” | “经过” } { <name of landmark feature 1>< name of landmark feature 2>… }

The rule defines the clause pattern which you will say normally when you arrive at several landmark features by the path. In the rule, landmark features are selected with method in section “IV.C”. A Chinese instance of the rule is “途中您可能看见建设银行”, expressed in English as “in the way you might see Construction Bank of China”.

D. Grouping Rule of Syntagms

Several clauses of PNLD will be constructed using the rules in section “IV.A”, “IV.B” and “IV.C”. A sentence of PNLD can be constructed by assembling those clauses in a sequence. The describing is as follow:

<Describing rule of path>::=<describing rule of start node>[{ <describing rules of non-direction-changed nodes>|<describing rules of direction-changed nodes> } …]<describing rule of end node>

The Controlling Knowledge of DPNL

The controlling knowledge includes not only the skilful knowledge about solving steps of some question which can guide user do something, but also the knowledge decides which action will be chosen when many actions are implementing at the same time^[5]. Some algorithms are used to realize corresponding functions in the process of DPNL, such as naming nodes of path, determining direction and selecting landmark features. The algorithms are the controlling knowledge. In this paper, programs are used to describe the controlling knowledge.

A. Naming Nodes

In daily life, nodes of road have not names normally. Therefore, nodes will need to be named firstly before constructing clause of PNLD of each node. There is a habit of naming node in daily life such as crossing of Ruida road and Kexue Avenue (expressed in Chinese as “瑞达路科学大道交叉口”) and t-shaped crossing (expressed in Chinese as “丁字路口”). The paper uses two elements as parameters of naming road nodes. One is NodeArcNum (ab, AN, Number of arcs correspond to node); another is name number of different arc which related to current node (DRoadNameNum, ab, N).

The controlling knowledge on naming nodes includes:

If AN equal to 1, then current node is the start node or the end node. Name of node should not be generated.

If AN equal to 2 or (AN equal to 4&N equal to 2), then current node is a crossing of two roads. The name of current node is named as “crossing of Road 1 and Road 2” expressed in Chinese as “路 1 与路 2 交叉路口”. Road 1 and Road 2 are different name of two roads.

If AN equal to 3, then current node is a crossing of three roads. The name of node is named as “trifurcate crossing” expressed in Chinese is “三叉路口”. Furthermore, if one road is vertical to another two roads at the same time, the name of node can be named as “t-shaped crossing” expressed in Chinese as “丁字路口”.

In other cases, current node must be the crossing of several roads. According to daily habits, they will be named as the “crossing” expressed in Chinese as “交叉路口”.

B. Determining Direction

Direction is important information of guiding of PNLD and also ensures the accuracy of PNLD. Direction can be divided absolute direction and relative direction. When current node is whether the start node or direction-changed node, judging of direction should be carried out. Absolute direction words in Chinese includes “东”(east), “南”(south), “西”(west), “北”(north), “东南”(southeast), “西南”(southwest), “东北”(northeast), “西北”(northwest), and so on. Absolute direction should be determined by longitude and latitude. Relative direction word in Chinese includes “前”(forward), “后”(behind), “左”(left), “右”(right), “左前方”(left-forward), “右前方”(right-forward), “左后方”(left-behind), “右后方”(right-behind), and so on. Relative direction could be determined using observer as the center and using polar coordinate system.

According to describing habit of GIS path in Chinese, the paper describes the path by a combinative method of using absolute direction and relative direction. In other words, describing direction of fist arc of road uses absolute direction; describing direction of residual arc uses relative direction.

The controlling methods include two parts. One part is determining direction of path by using the angle between first arc and Northern when describing first arc of path. The other is determining direction of path by using polar angle (θ), illustrated in Fig.1 in polar coordinate system which origin is current node when describing arcs of path except first arc.

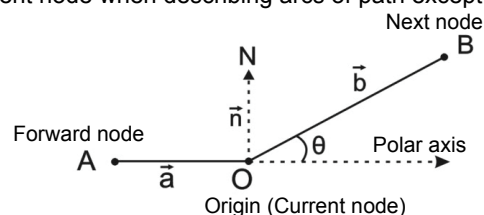


Fig.1. Model of Polar Coordinate of Path

Table 1. Corresponding Relationship between Polar Angle and Direction

Range of Polar Angle	Direction
-22.5°~22.5°	Forward
22.5°~67.5°	Left-forward
67.5°~112.5°	Left
112.5°~157.5°	Left-behind
-67.5°~-22.5°	Right-forward
-112.5°~-67.5°	Right
-157.5°~-112.5°	Right-behind
157.5°~180° or -180°~-157.5°	Behind

According to daily habit of people, the corresponding relationship between polar angle and direction in the paper is set in Table.1.

C. Selecting Landmark Features

In process of describing of PNLD, a few landmark features beside road not only identify and guide the correct road but also enrich the describing sentence. Number of landmark features used in sentence is not so many. Otherwise, not only the describing sentence is so long, but also the main information of path could not be understood easily. Landmark features are collected and stored in one map layer. In the collecting process, number of landmark features is controlled in reason and landmark features are associated with arcs of road with which not only the selecting process is simplified but also the symbolic effect of landmark features could be insured.

In selecting process of landmark features, the properties of LmFeatureNum(Number of landmark features) and LmFeatureID(ID collection of landmark features) are used. If LmFeatureNum>0, the algorithm of PNLD acquires the names of landmark feature from landmark features' name base. Then, describing clause could be constructed using those names of landmark feature related to current arc of road. If there are several landmark features related one arc, the sequence of landmark features is confirm by comparing coordinates of those features and then the names of landmark features could be organized in the sequence.

Testing Describing Knowledge of DPNL

Based on summarizing all kind of describing knowledge of DPNL, the paper builds describing knowledge base which includes several text files stored describing knowledge. Using part road data of Zhengzhou city, a prototype system of DPNL is developed using algorithm of PNLD based on continuous polar coordinate^[7] by C# program language. The describing knowledge of DPNL is test using route sketch describing function and natural language (Chinese) describing function in prototype system. According to the classified method of knowledge which classification includes the fact knowledge, the rules knowledge and the controlling knowledge, the paper analyses and concludes all kind of description knowledge during the describing process of path, and builds the describing knowledge base of path. Finally, the paper realizes the prototype describing system by which the describing knowledge is test.

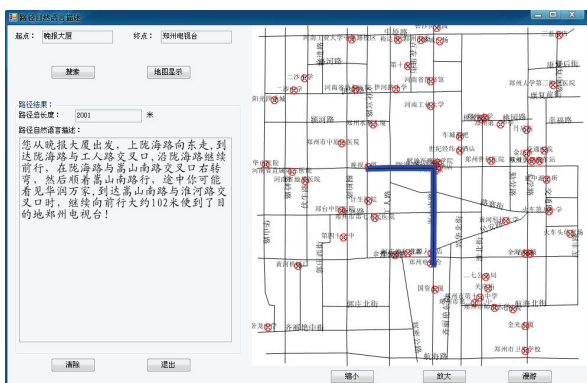


Fig.2. Describing of GIS Path Based on Route Sketch and Chinese Sentence

In Fig.2, the describing result of path from “晚报大厦” (The Evening Newspaper Office Building) to “郑州电视

台”(Zhengzhou TV) is displayed. In route sketch, the line in bold expresses the shortest path; in the box of describing result, the Chinese sentence describes the shortest path in Chinese. The sentence expressed in Chinese is “您从晚报大厦出发, 上陇海路向东走, 到达陇海路与工人路交叉口, 沿陇海路继续前行, 在陇海路与嵩山南路交叉口右转弯, 然后顺着嵩山南路行, 途中您将看见华润万家, 到达嵩山南路与淮河路交叉口时, 继续向前行大约 102 米便到了目的地郑州电视台。”, expressed in English as “You can start from the Evening Newspaper Office Building; go along with Longhai Road towards east. When you arrive at the crossing of Longhai Road and Gongren Road, move on along Longhai Road. You should turn right at the crossing of Longhai Road and Songshan Road, and go along with Songshan South Road, in the way you will go by Vanguard supermarket. When arriving at the crossing of Songshan South Road and Huaihe Road, you can walk on about 102 meters to arrive at the terminus, Zhengzhou TV.”.

In query result, describing result of DPNL includes description of the absolute position (East), the relative position (left), landmark features beside road (Vanguard Supermarket) and the proper distance. It is improved that the describing sentence of query result built by implementing the describing knowledge of DPNL concluded in the paper overcomes describing rigidly and understanding difficultly without enough auxiliary information in other natural language description methods. The describing method using DPNL is accurate, simply, vivid and accords with the describing habit of people in daily life.

Supported by the Fund Project of Zhengzhou Institute of Surveying and Mapping (Grant No. Y1004).

REFERENCES

- [1] Hoepfner W, Carstensen M, Rhein U. Die Interdependenz von Such- und Beschreibungsprozessen. Repräsentation und Verarbeitung räumlichen Wissens,1990,(245):221-234.
- [2] LIU Yu, GAO Yong, LIN Bao-jia, WU Lun. Research on GIS Path Reconstruction Based on Constrained Chinese Language. Journal of Remote Sensing,2004,8(4):323-330.
- [3] ZHANG Xueying, LV Guonian, HUAN Jian. Approach to Route Description in Chinese Natural Language for Electronic Maps. Geo-Information Science, 2008,10 (6) :757-762.
- [4] GAO Hongshen. Decision Support System: Theory and Methods(4rd). Beijing:Tsinghua University Press, 2009.
- [5] CHENG Xianyi, LIU Yisong, YAN Li. Knowledge engineering agent-oriented,Beijing: Science Press, 2008.
- [6] LI Bin. Research on Interpreting Mechanism of Natural Spatial Query Language.PLA Information Engineering University,2009:83-84.
- [7] CHANG Dehai, LI Bin, ZHANG Min The Chinese Describing of GIS Path Based on Continuous Polar Coordinate,Journal of Geomatics Science and Technology.2011. 28(4):307-312.

Authors: PhD students. Dehai CHANG, School of Surveying and Land Information Engineering, Henan Polytechnic University, No.2001, Century Avenue,Jiaozuo, Henan Province,China, E-mail:changdehai@126.com; PhD. Bin LI, Zhengzhou Institute of Surveying and Mapping, No.66, Longhaizhong Road, Zhengzhou, Henan Province,China, E-mail: zb35libin@sina.com; prof.Qingshan WANG, Zhengzhou Institute of Surveying and Mapping, No.66, Longhaizhong Road, Zhengzhou, Henan Province,China.