The Keyword Search Using Thesaurus Concept in Geographic Information Systems

Takashi Yamauchi, Dongshik Kang and Hayao Miyagi
Department of Information Engineering, Faculty of Engineering,
University of the Ryukyus,
1 Senbaru, Nishihara, Okinawa 903-0213, Japan
e-mail : takashi@net.ie.u-ryukyu.ac.jp
Tel.: +81-98-895-2221(ext. 2519), Fax.: +81-98-895-8717

Abstract: In this study, it enabled to perform keyword search on Geographic Information Systems (GIS) more flexibly by taking in the concept of geographic thesaurus in order to make the keyword search that was more efficiently. The search procedure depends on the history when user information is included. This study provides a system for keyword searching as well as check the validity of the system. Furthermore, by establishing this reference method, the search that took in the query of a user flexibly is attained rather than a chosen type reference system.

1. Introduction

The spacial data analysis using the Geographic Information Systems (GIS) attracts a great deal of attention with highly advancement of computer technology in recent years. It is possible to search a required data since the geospacial data in GIS contains such as position information, topological information and attribute information. Now, it is difficult to use keyword search because it depends on input keyword that corresponds to the attribute information of the geospacial data. Moreover, the system reflecting user's feature and tendency is required because different user often differs from search tendency.

Here, we propose the keyword search system by using the concept of the thesaurus in GIS. Therefore, it is possible to support the way of thinking of user for a search query. Moreover, personalization of search tendency is obtained by setting up and referring to the past history when the user uses this search engine.

2. GIS and Thesaurus Concept

2.1 GIS

Geographic Information Systems (GIS) is systems which can perform spacial reference, analysis, display etc. by managing the spacial data obtained in different scale, time and dataform. GIS attract attention quickly with the reduction of cost in computer processing environment[1].

The treated data by GIS connects the position reference data which includes earth surface, the ground and underground etc. and attribute data, such as nature, society and economy etc. Generally this data is called geographic spacial data. Futhermore, position reference data is called topological data. Topological data is expressed the object on the earth, such as the latitude and longitude, figure.

Attribute data is stored in database which includes figure data, numerical value and character. For example, if it is as attribute data on a building, it will be stored the address of the building, telephone number, owner, etc. as attribute information.

2.2 Thesaurus Concept

The thesaurus hierarchy is classified into words and concepts[2]. The thesaurus has an equivalent relation between words, such as a synonym, a wide sense word and a narrow sense word, etc. For example, correlation of "computer" with thesaurus to extract related word(s) is shown in Figure 1. We can discover easily the related word by using the thesaurus.
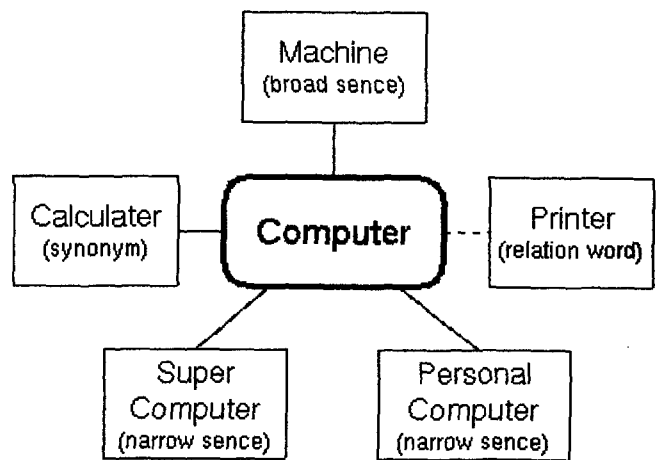

Figure 1. Correlation with thesaurus of "computer"

Moreover, the relation of each word is acyclic graph that is compared to words as node and relations as link. Therefore, it becomes possible to process easily the search of arbitrary words by various calculations. However, real thesaurus builds actually, it designed in many cases focusing on the related word collected from the specific field because the limitation of recorded words, detailed construction, etc. In case of geographic field, relation words are mainly considered to be stored:

● Region name
A formal name, a common name, the located cities, towns and villages name, etc.

● Building name
A formal name, a common name, a contained tenant name, etc.

- Road name
  A formal name, a road number, a common name etc.

- The kind of action
  Drive, a travel, reading etc.

Moreover, thesaurus constructions change is needed flexible because the position of the word may be changed with the added word(s). Research on thesaurus is also studied actively.

2.3 Present Keyword Search in GIS

Generally, Internet GIS is being used widely[3]. Internet GIS is used for reference of the destination on a map in many cases. Mapion which can search the information on landmarks is known as one of the Internet GIS[4].

In Mapion, there are the two reference methods namely the multiple-choice search system and the keyword search system. Both of the systems mentioned to the following features:

- Multiple-choice search system
  Search system has been prepared a selection item beforehand by the system designer, and a user chooses an item to investigate from a selection item. The number of selection item that can use search system is small. For example, the cities, towns and villages in an area are suitable for the search system since the number is decided.

- Keyword search system
  A user inputs into input form the query, and obtains from the item that is in agreement with a keyword. When the number of a selection item is not a constant or there are quite a many items, keyword search system is very effective. For example, many kinds of sightseeing article on souvenir shop in each country. When a user searches one kind of query, the keyword search for direct input is considered easily rather than it chooses the item from many sightseeing names of articles.

When using search system, if the user query is not included in the selection item, and a selection item is ambiguous as for the user, it does not fill the reference query. When use keyword search system, it returns a result only when in agreement with the keyword which compared with the database according to the user input. No output will be obtained when input keyword does not matched in database. Then user has to think a new keyword newly. So, it can be said that the keyword search system is not enough effective.

3. Method of Keyword Search System Using Thesaurus Concept

This section describes the methods of keyword search system in this research. Furthermore, the flow in this system is described in detail.

3.1 The Flow of Keyword Search System

In this study, two new databases namely the geographic thesaurus database and the history database have been created in order to build the keyword search system. The procedure of this keyword search system is briefly shown in Figure 2.
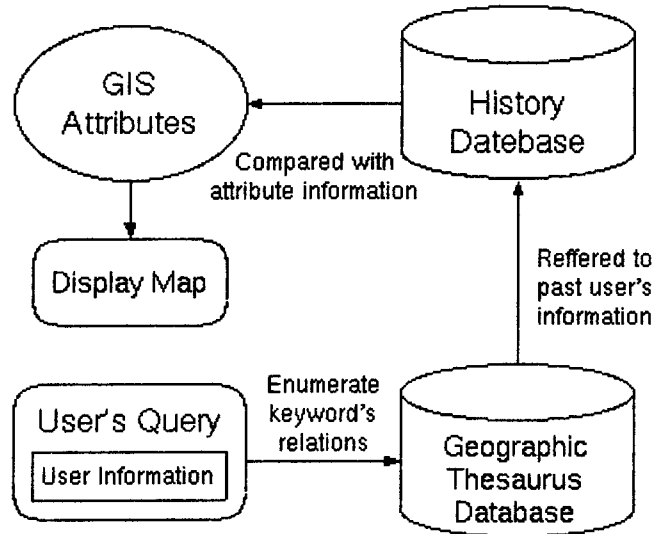

Figure 2. Search procedure of keyword search system

3.2 User Information and User Query

When using the system, each user inputs the personal information such as "user name", "age", and "sex" at later. The inputted data is stored in history database. User inputted personal information could be retrieved by using "user name" only.

3.3 Geographic Thesaurus Database

The geographic thesaurus database is constructed by using existing thesaurus dictionary in this study.

The geographic thesaurus database is registered only nouns such as a company name, a goods name, etc. Each word applicable to a synonym, a wide sense word, a narrow sense word, etc. is stored in the geographic thesaurus database.

The keyword inputted by the user is analyzed and compares with the words registered into the geographic thesaurus database. Then, the geographic thesaurus database extracts the words relevant to the keyword.

3.4 History Database

The history database is stored user's personal information that includes user name, user's age, user's sex, inputted keyword by user, the selected word during feature selection, and number of times of selection. Each element of the history database is shown in Table 1.

A value attachment procedure executes to attach a value for each word. Determination of value for each word can be obtained by the following procedure:

576

Table 1. The example of a history database

User	Age	Sex	Input Word	Section Word	Times
Arisa	18	F	computer	printer	5
Hanako	28	F	computer	DVD-ROM	1
Ichiro	25	M	computer	keyboard	7
Takashi	23	M	computer	Electrical store	3
Takashi	23	M	computer	Computer shop	7

1) It extracts the agreement item that an inputted keyword by user and "input word" in the history database.

2) The value attachment is performed in the "selection word" which attends the extracted input word group in order of difference with the existence of "user name", the "number of times", user's "age" difference, and "sex."

3) When the same "selection word" exists, the line with a low priority deletes.

In case of new user and "user name" of the history database don't corresponds with user's name using now, the value attachment is performed from "number of times". Moreover, if no agreement item is found in the history database, then it creates output by correlate the words in the geographic thesaurus database. When a user is "Takashi", this procedure performs priority attachment on Table 1, the result will be obtained Table 2.

Table 2. Result of sorted history database

User	Age	Sex	Input Word	Section Word	Times
Takashi	23	M	computer	Computer shop	7
Takashi	23	M	computer	Electrical store	3
Ichiro	25	M	computer	keyboard	7
Arisa	18	F	computer	printer	5
Hanako	28	F	computer	DVD-ROM	1

The related word group in which priority attachment was performed using the past history in a history database compares with attribute information.

3.5 Collation with GIS Attribute Information

The related group of words to which value was attached has shown user by Graphical User Interface (GUI). The word group has been arranged according to the value that word. Moreover an output relation words group is shown, the feature that attribute information on geospacial data corresponds with an input keyword by user is displayed. User can select anyone arbitrary words from output relation word group, and can aim at agreement with the feature

demanded recursively using the new output word group derived from the selection word.

If the feature of attribute information corresponds with the keyword is selected then the data including user personal information (user name, user's age, user's sex), input word and selection word are stored in the history database. The stored data can be adapted for use for next time.

4. Experiment and Results

The system proposed by this research is actually built and checked. In this study, the keyword search system using the thesaurus concept has been built using Ryutan[5].

The example of an experiment is shown in Figure 3. Figure 3 is the initial screen of search. A result is displayed when a user inputs a keyword into the input textbox under upper left of "Keyword".
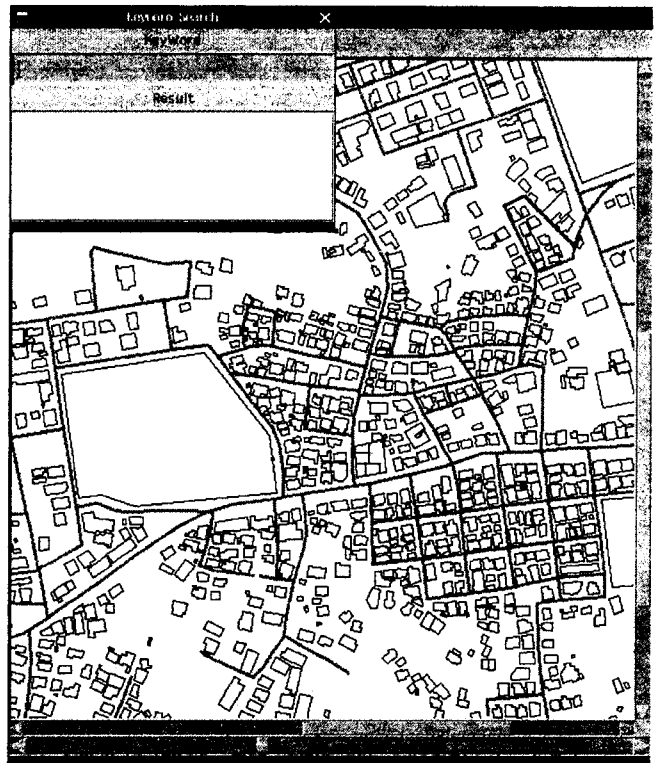

Figure 3. The initial screen of search form

When the keyword is not stored in the geographic thesaurus database, it massages there was no corresponding word. When the keyword inputted user is stored in the geographic thesaurus database, it output to a column like Figure 4 under "Result" whose word relevant to keyword. The related word group output, a word with a high priority is displayed sequentially at the top using the history database. Moreover, it colors in the building attribute information and whose results correspond on a map.

Furthermore, in double-clicking by choosing one from the result word group is displayed, the selected word turns in a new keyword. In this case, new related word displays

Figure 4. The display of the search result inputting "computer" into the input textbox
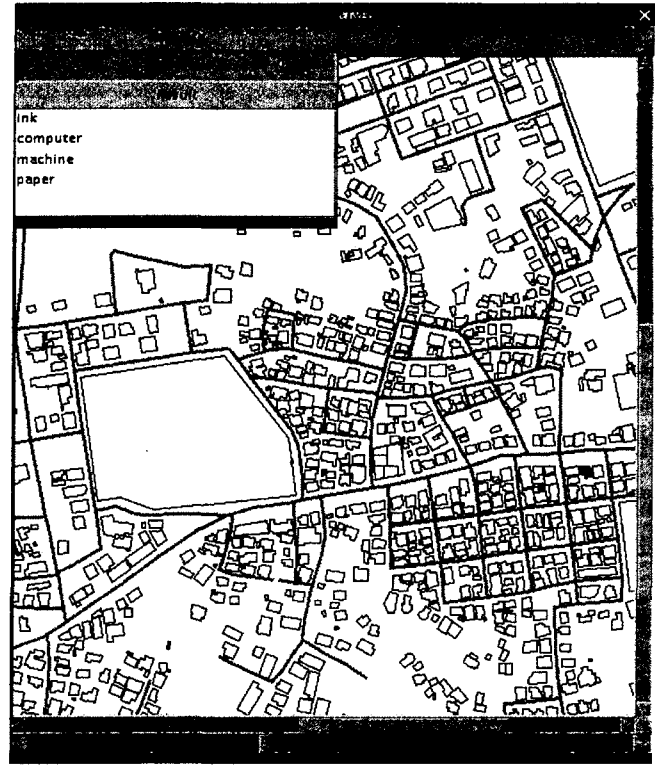

Figure 5. The display of the search result at the time of input "computer" and choosing "printer" as searching a new reference query

on the column under "Result." An example is shown in Figure 5. This can be performed continuously. It is possible for a user to look for the related word visually related by easy operation, and to look for the building and a road that fills a reference query.

5. Conclusions

In this study, the concept of the keyword search system that used the geographic thesaurus on GIS has been proposed. Moreover, the technique reflecting the search tendency for each user is proposed by performing the value attachment of related words using the history database.

In the future, completion of the geographic thesaurus database, realization of keyword search in Japanese etc. will examine.

References

[1] D. J Maguire, M. F Goodchild and D. W Rhind, *Geographical Information Systems: principles and applications*, Longman Group UK, 1991

[2] J. Aitchison, A. Gilchrist, *Thesaurus Construction: A Practical Manual, 2nd Edition*, Aslib, 1987.

[3] B. Plewe, *GIS Online: Information retrieval, Mapping, and the Internet*, Delmar Publishing, 1997.

[4] Mapion [Information Map Comunication], http://www.mapion.co.jp/

[5] Y. Nie, M. Nakamura, H. Miyagi and K. Onaga, "A design and implementation of spacial data transfer system with object oriented technology -ryutan", *Theory and Applications of GIS*, Vol.8 No.2, pp.51-59, 2000.