

# Two-phase Query Modification using Semantic Relations based on Ontologies

**Kaoru Hiramatsu Jun-ichi Akahani Tetsuji Satoh**

NTT Communication Science Laboratories  
Nippon Telegraph and Telephone Corporation  
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237 Japan

## Abstract

We propose two-phase query modification using ontologies for geographic information navigators. The query modification is divided into two phases: 1) revising ambiguous conditions into appropriate ones automatically to obtain an adequate number of search results, and 2) providing next available options to enable users to jump to advanced and associated topics. Both phases are processed using semantic relations based on ontologies and conduct users to query evolution interactively. While processing, our prototype system outputs search results in various styles, such as digital maps and tables, so that it enables users to have a good understanding of the relational structures among the search results. In this paper, we explain an outline of the prototype system and examine the effect of the query modification through an example.

## 1 Introduction

Current search engines have become powerful tools on the Internet, but they are too strict and too inflexible to handle queries that are entered by users. For example, in querying geographic information, there are some cases where users cannot get the intended result because of the difficulty in specifying geographic conditions. If a user does not have sufficient knowledge of the targets, results that include unrelated areas and objects will be listed. If a user knows the exact names of the intended targets, only pinpoint results will be output. These occasions are not restricted to geographic information and can occur in various domains. Even though some search engines provide a spelling correction function, the current adaptation capability to deal with a user's query is still limited.

Moreover, after the initial query, it is hard to evolve the original query into advanced and associated topics because of the difficulty in understanding the characteristics of the search result and its circumstances. Since many search engines only display the search result in a list form, neighbourhoods that exist close to the search result of the original query are rarely discovered by the users and the systems.

In order to avoid these inconveniences, it is efficient to utilize semantic relations among the information. For instance,

synonyms, hypernyms, hyponyms, and siblings found in thesauri and WordNet[WordNet] enable the system to replace terms in the original query with more effective ones and conduct the user to advanced and associated topics. In querying a specific spot, gazetteers (e.g., [Hill, 2000]) and digital maps (e.g., [Ambite *et al.*, 2002]) are also helpful. According to the number of search results, the original geographic conditions are expanded into a wider area or moved to surroundings along the geographic relations in the real world. Even if the user does not know the exact name of the spot, combinations of partial and incomplete conditions also have the potential for leading to an adequate number of search results.

In this paper, we employ the query modification process into our prototype system of geographic information navigators that retrieve and display a set of Web pages with relationships on the Internet and in the real world. For utilizing this query modification process in our system, we divide the process into two phases:

1. Revising ambiguous conditions into appropriate ones for getting an adequate number of search results, and
2. Providing next available options to enable users to jump to advanced and associated topics.

The first phase is processed automatically during query processing to avoid outputting a zero search result or a huge result list. The second phase is invoked after query processing and requires the user's selection. Both phases are processed in accordance with semantic relations derived from meta data, thesauri, and gazetteers that are based on ontologies and conduct users to interactive query evolution.

In the following, we first explain an outline of our prototype system that consists of a user interface module, a query processing module, and a database management module. The prototype system enables the user to retrieve a set of pointers to Web pages that are connected by hyperlinks and implicit geographic relations between Web pages. Then, we propose a two-phase query modification and demonstrate how an entered query is modified using an example in Kyoto, Japan. Finally, we discuss how we should utilize semantic relations among contents and ontologies.

## 2 Geographic Information Navigator

To enable users to have a good understanding of the relational structures among search results, we have refined our

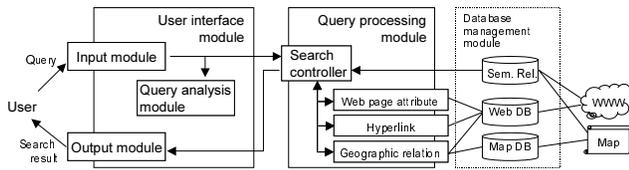


Figure 1: System architecture

map-based Web page search system [Hiramatsu *et al.*, 2000] as a geographic information navigator by integrating functions of interactive search and easy query modification. The user first enters a natural sentence query or fills a query form, and then gets a search result that includes not only a set of pointers to Web pages but also relational structures among them. According to the search result, the system also provides advanced and associative queries that are evolved from the original query. Throughout this interaction loop, we aim to deepen and broaden the user's knowledge about the target.

The prototype system consists of a user interface module, a query processing module, and a database management module (Figure 1). These modules communicate with each other via a network. Data sources of the prototype system are Web pages on the Internet and detailed digital maps. The Web pages are collected by crawling and analyzed to extract some features (such as URL, title, and keywords of contents) as meta data. Furthermore, if phrases in the Web page match an object's name and address on the digital maps, we also add the geographic coordinate of the point of the object to the meta data. These meta data are stored on the database management module.

The query processing module is based on an augmented Web space [Hiramatsu and Ishida, 2001] that enables it to search a set of pointers to Web pages with relational structures by specifying conditions of the Web page attributes, hyperlinks, and implicit geographic relations between the Web pages. In the augmented Web space, geographical generic links between Web pages are created dynamically according to implicit geographic relations, such as distance and directions, between objects that are described in the Web pages. Although the augmented Web space is designed to utilize only geographic relations originally, it is easy to integrate many types of semantic relations derived from the meta data, thesauri, and various aspects in the same manner.

The user interface module is the front end of the prototype system that handles the input of queries and output of search results (Figure 2). The input module accepts a query either in natural language style or in form style. In the natural language style, the user can enter a query in a natural sentence. The query analysis module parses it and translated it into a query in the extended SQL for the augmented Web space. The translated query is transferred to the query processing module and also reflected in the form style interface. Each field of the form corresponds with conditions in a query written in the extended SQL so that the user can edit the translated query again. Of course, the user can enter a new query to the form directly.

The search result that is returned from the query process-

ing module includes a set of pointers to Web pages with relational structures among them. The output module of the user interface module displays them using various views, such as digital maps, lists, tables, and 3D tree views (Figure 3). For example, link relations among the search results are displayed in the 3D tree view and the pointers of the Web pages that have geographic coordinates as meta data are superimposed on the digital maps. Moreover, if the user selects a search result in one of these views, corresponding items in other views will be focused on synchronously. This synchronization aims to help users obtain a good understanding of the relational structures among the search results.

### 3 Two-phase Query Modification

We propose two-phase query modification using ontologies for the geographic information navigators. This function is applied to our prototype system for conducting users to query evolution interactively. We suppose it will help users deepen and broaden their knowledge of city information.

The query modification process is divided into two phases, which we call "two-phase query modification" (Figure 4). The phases are processed by the user interface module and the query processing module cooperatively.

In the first phase, the query processing module prefetches search results of each term condition in the query and identifies ambiguous term conditions in accordance with the number of search results. The module then revises them into appropriate ones automatically using semantic relations. For example, if a keyword in the query is too ambiguous and a huge amount of search results are output, the keyword will be replaced with a more specific term along with word relations in the thesaurus. On the other hand, if a keyword results in a zero search result, then the keyword will be replaced with a hypernym or a category name. After revising the ambiguous term conditions, the module modifies inadequate link conditions in the query. If a geographic condition between Web pages is too narrow, in that an adequate result cannot be produced, the condition will be expanded into a wider area according to the characteristics of the database. The module backtracks and revises the conditions repeatedly until an adequate search result is obtained.

In the second phase, next available options are provided to the users to enable them to advance the first query to a more advanced one and turn the original target into associated topics. Since the original search result has a tree structure, the next available options are developed from root and leaf nodes of the search result and from characteristics of links between Web pages according to semantic relations. The user is prompted to choose one of the options explicitly, while the query is modified automatically in the first phase. The current version of the prototype system does not evaluate the fitness of the options using some kind of utility function, so that all the developed options will be provided to the users without priority order.

### 4 Experimental Example

To experiment on our prototype system with practical databases, we collected Web pages that are related to ob-

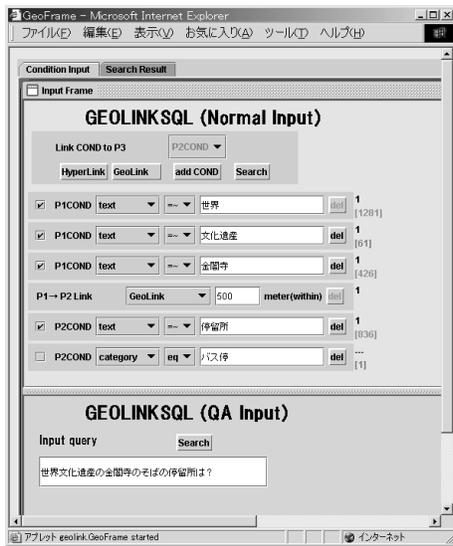


Figure 2: Input frame of the user interface

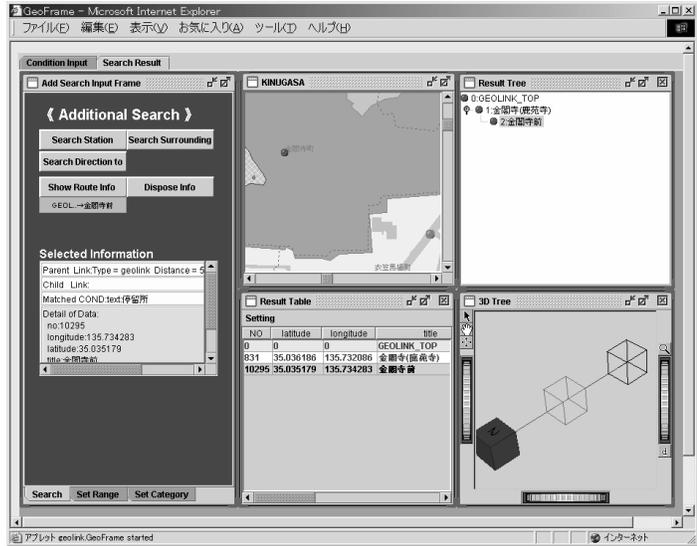


Figure 3: Output frame of the user interface

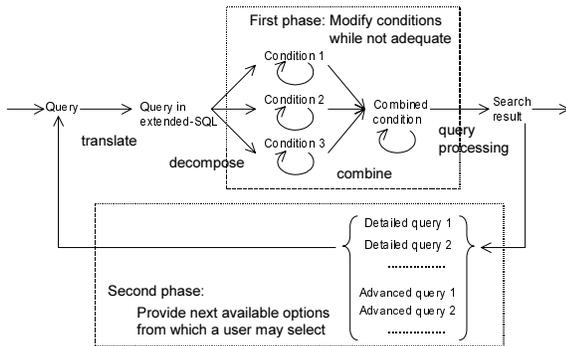


Figure 4: Two-phase query modification

```

SELECT p2.url
FROM p1,link,p2
  p1.keyword =~ 'Kinkakuji'
  p1.keyword =~ 'World Heritage'
  link AS distance(p1,p2) < 250m
  p2.keyword =~ 'bus stop'

```

Figure 5: Translated query in extended SQL

jects in Kyoto, Japan. We extracted attributes of the Web pages such as URL, title, category, keywords, incoming links, outgoing links, and geographic coordinates the Web pages may have and put them into databases of the prototype system. Furthermore, hierarchies among categories and words based on Nihongo Goi Taikei (a Japanese thesaurus) [NTT-CS, 1997], digital maps of Kyoto, gazetteers extracted from the digital maps are also stored in the databases as semantic relations. Because the contents in the databases are mainly in Japanese, we will use an example query in Japanese and explain how the processing is carried out in English in this paper.

The example query is “find bus stops near Kinkakuji temple, which is one of World Heritage” that is shown in Figure 2. A user inputs the example query into the field in the lower part of the frame. Then, the user interface module analyzes the query and translates it into a query in the extended SQL for the augmented Web space (Figure 5). The translated query is also displayed in the upper part of the same frame so that the user can edit conditions of the query using this form after

getting the first search result.

After translation, the first phase of the query modification process is invoked. The number of matching results for each term condition in the translated query is estimated by the query processing module. In the example query, there is no term condition where the estimated result is zero or beyond the preset thresholds. If there is such a condition, then it will be replaced with more adequate term conditions using semantic relations.

Next, matching results are retrieved from the databases and combined into the final search result in accordance with the link conditions. The example query succeeded in retrieving a sufficient result; however, if the combined result becomes invalid even though each term condition is valid, the system backtracks to the term replacement process and re-combines the results until an adequate result is obtained. Also, if a link condition based on geographic distance is invalid, the system extends or shortens the distance in order to get the adequate final result.

The search result of the example query is shown in Figure 3. On the map frame of the output frames, the upper left icon is an anchor of the link to the Web page of “Kinkakuji temple” and represents a location of it. In the same way, the lower right icon is an anchor of the Web page of the nearest bus stop and represents a location of it. The attributes are listed on the table frame and their relations are displayed on

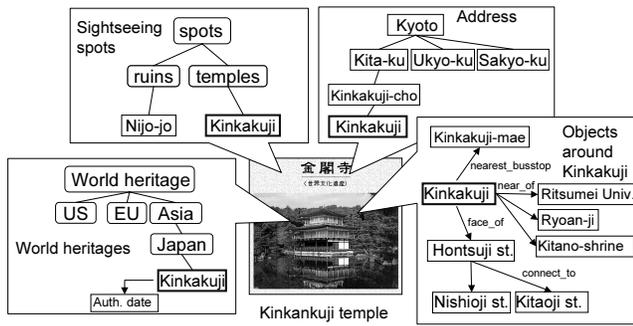


Figure 6: Semantic relations among “Kinkakuji”

the 2D/3D tree view frame.

In the second phase, pointers of next available options are displayed on the output frame. In this example query, the user receives pointers of the advanced and associated topics, such as Web pages that are linked from the Web page of Kinkakuji temple and the bus stop, all of the information or the filtered information surrounding them on the map, and a route information between them. Furthermore, the user can enhance the query recursively according to the search result of the modified query. Although this query modification is based on semantic relations such as the gazetteers in the digital maps and the hierarchies in the thesauri, more semantic relations will help the prototype system develop more options.

## 5 Discussion

The two-phase query modification we proposed in this paper helps the users enhance the original query. The revision of inadequate conditions is a small step in query modification. It avoids annoying the users with keyword adjustment to get adequate search results automatically. On the other hand, the next available options are regarded as chances to jump from the original query to advanced and associated topics.

Although we prepared the semantic relations for the experiment by ourselves, many other relations among objects are dormant. For example, in Figure 6, instances of Kinkakuji temple appeared in four semantic relations and are related to other instances. They explain that Kinkakuji temple belongs to the temple category in sightseeing spots, that it is located in Kinkakuji-cho, Kita-ku, Kyoto, and so on. These semantic relations have the potential to augment the query modification.

On the semantic Web, it is expected that the system can import such semantic relations from the Internet. The World Wide Web Consortium [W3C] is developing some languages such as RDF and OWL for describing the semantic relations and meta data. This means that everyone can publish semantic relations so that formal hierarchies and user-developed relations can be created. Moreover, there are several ways to derive Web page relations using semantic relations. The geographical generic links we proposed in the augmented Web space are created according to geographic relations, such as distance, direction, and the inclusion relation, using the geographic relational operation of geographic information sys-

tems (GIS). The hierarchies of categories and keywords can be applied not only to modifying the query but also to similarity evaluation and Web page clustering. Most of these relations are transitive and functional, so the system should take care how far it modifies and extends the original query using the relations.

## 6 Conclusion

In this paper, we proposed two-phase query modification for geographic information navigators and explained how it works using practical databases in Kyoto, Japan. As discussed in Section 5, the semantic relations will have heterogeneity on the semantic Web, so that the system should evaluate accuracy, reliability, and suitability of the semantic relations carefully before employing them. We are planning to integrate such meta level strategies into our prototype system and evaluate them from the viewpoints of system performance and users’ knowledge acquisition in future work.

## Acknowledgement

Thanks are due to content holders for the permission to use their Web pages concerned with Kyoto, and to Yuji Nagato of NTT Comware Corporation for his great contributions to the prototype system.

## References

- [Ambite *et al.*, 2002] Jose Luis Ambite, Greg Barish, Craig A. Knoblock, Maria Muslea, Jean Oh, and Steven Minton. Getting from Here to There: Interactive Planning and Agent Execution for Optimizing Travel. In *Proceedings of the Fourteenth Conference on Innovative Applications of Artificial Intelligence (IAAI-2002)*, pages 862–869, 2002.
- [Hill, 2000] Linda Hill. Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints. In *Research and Advanced Technology for Digital Libraries: Proceedings of the 4th European Conference, ECDL2000*, pages 280–290, 2000.
- [Hiramatsu and Ishida, 2001] Kaoru Hiramatsu and Toru Ishida. An Augmented Web Space for Digital Cities. In *The 2001 Symposium on Application and the Internet (SAINT2001)*, pages 105–112, 2001.
- [Hiramatsu *et al.*, 2000] Kaoru Hiramatsu, Kenji Kobayashi, Ben Benjamin, Toru Ishida, and Jun-ichi Akahani. Map-based User Interface for Digital City Kyoto. In *INET2000 The Internet Global Summit*, 2000. [http://www.isoc.org/inet2000/cdproceedings/4c/4c\\_1.htm](http://www.isoc.org/inet2000/cdproceedings/4c/4c_1.htm).
- [NTT-CS, 1997] NTT Communication Science Laboratories. *Nihongo Goi Taikai*. Iwanami Shoten, 1997.
- [W3C] World Wide Web Consortium. <http://www.w3c.org/>
- [WordNet] WordNet. <http://www.cogsci.princeton.edu/~wn/>