INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Coping with geographical relationships in Web searches

Lin Tzy Li *Ricardo da Silva Torres*

Technical Report - IC-10-04 - Relatório Técnico

January - 2010 - Janeiro

Coping with geographical relationships in Web searches

Lin Tzy Li* Ricardo da Silva Torres†

Abstract

This work describes ongoing research which aims to evaluate the limitations of existing tools to support geographic information retrieval on the Web. More specifically, we are interested in assessing how well queries involving geographic relationships can be specified and executed using existing Web tools. Our evaluation considered three different aspects: the relevance of the results and the time and complexity of the steps necessary to obtain these results. More than thirty Web users performed five tasks, considering five scenarios. Experiment results demonstrate that existing Web tools are not enough integrated to meet user needs regarding the specification and execution of queries involving spatial relationships.

1 Introduction

Geographic information refers to the existence of an attribute which is related to a localization on Earth, for example a geographic coordinate or some relation to some other object whose geographic location is known. It might be a fully complete address or even a single reference.

The geographic information is present in people's daily life, thus it is not surprising that there is great amount of information on the Web about geographical entities and great interest in localizing them on maps. Tools like Google Maps and Google Earth are very popular and they partially meet the needs of Web users for geospatial information.

On the other hand, the conventional search services are based on keywords matching and do not consider that those words might represent geographical entities which spatially relates to each other. Even though these relationships have not been explicitly used in a query, they are potentially relevant to users [18].

An example of query that most of the existing *Geographic Information Retrieval* systems do not support is: "Which are the Web pages of the cities which are neighbors of Campinas County". The difficulty of processing this kind of query relies on combining traditional queries executed on Web search mechanisms with spatial operators usually implemented in spatial databases.

This work aims to assess how well queries involving geographic relationships can be specified and executed using existing Web tools. Our evaluation considered three different

*Institute of Computing – University of Campinas (UNICAMP), 13081-970 Campinas, SP & Fundação CPqD, Campinas, SP.

†Institute of Computing – University of Campinas (UNICAMP), 13081-970 Campinas, SP.

aspects: the relevance of the results and the time and complexity of the steps necessary to obtain these results. More than thirty Web users performed five tasks in a experiment considering real scenarios. Our study does not intend to evaluate any specific tool. In fact, users were encouraged to use whatever Web tool they wanted to perform the proposed tasks.

The hypothesis of the experiment is that the processing of queries involving geographical relationships between geographical objects, using the Web existing tools, is complex, does not return enough relevant results, and is time-consuming. Although that might sound obvious, to the best of our knowledge, there is not study about Web tools that consider the execution of Web queries involving geographical relationship by potential users. This work intends to confirm those hypotheses, as well as to shed some light on how the users perform queries, and then deriving strategies to improve current search tools. Experiment results demonstrate that existing Web tools are not enough integrated to meet user needs regarding the specification and execution of queries involving spatial relationships.

Next section describes the conducted experiments while section 3 presents and discusses obtained the results. Section 4 presents our proposal to integrate current Web tools aiming to support queries involving geographic relationships. Section 5 discusses related work, while section 6 states our conclusions and future works.

2 Experiment

This section describes the methodology of the conducted experiment[1].

2.1 Methodology

An invitation letter and a set of questionnaires were sent by email to Web users with no profile restrictions. The users were encoraged to forward them to other users under their network. It did not mean to reach every kind of Web user, but whoever willing to participate to this experiment.

The users were invited to answer an opening questionnaire whose results allowed us to characterize their profiles. Next, they performed five contextualized tasks. At this point, users were asked to time the task execution from the beginning to the end, to describe the steps used to perform the task, to list the retrieved results, and finally to assess the satisfaction with the time spent and with the retrieved results. Finally, opportunity was given to the users to make suggestions or comments about the existing tools and other issues related to the experiment in a closing questionnaire.

Tasks: The tasks were designed to include different geographical relationships, from topological relation, as *touch* (neighbor) and *cross* [7], to metric and directional relations like *near* and *south*. At first a main questionnaire was designed with 8 tasks, but as they would demand too much time from the users, it was cut down to five representative tasks in an attempt to make it less time consuming and therefore attract as much voluntaries

[1]The execution of the experiments conducted in this research was approved by the Unicamp's Committee for ethics in research (CEP). Its is registered at the Brazilian National Commission of Ethics in Research (Comissão Nacional de Ética em Pesquisa - CONEP) by FR - 240244.

as possible to participate in this experiment. The proposed tasks and their contexts were given as follow:

Task 1: Search for Web pages of hotels which are out of city of Curitiba, but still in the Curitiba's metropolitan area (the region of interest is illustrated in Figure 1).

Task context: Despite your budget restriction, you are going to visit Curitiba. You know that hotels in Curitiba might be more expensive, so you would like to search for hotels in the cities surrounding Curitiba where you can stay and still visit Curitiba.

Figure 1: Task 1: Curitiba city and and the cities of Curitiba's metropolitan area.

Task 2: Search for Web pages of cities in the neighborhood of Curitiba (see Figure 2).

Task context: You live in Curitiba and you would like to apply for public contests to work in Curitiba or in Curitiba's metropolitan area. Therefore you would like to reach city's Web pages to look for announcement for open public positions.

Task 3: Search for Web pages of concessionaires of highways that cross Tietê River (see Figure 3).

Task context: In addition to traveling throughout São Paulo state, you are interested in taking the chance to research about towns and cities near Tietê River. It would be interesting to find highways concessionaire's Web pages that manage the highways which cross the river in order to contact them and to propose some kind of sponsorship or join work.

Task 4: Search for Web pages of Barcelona's hotels which are near subway stations.

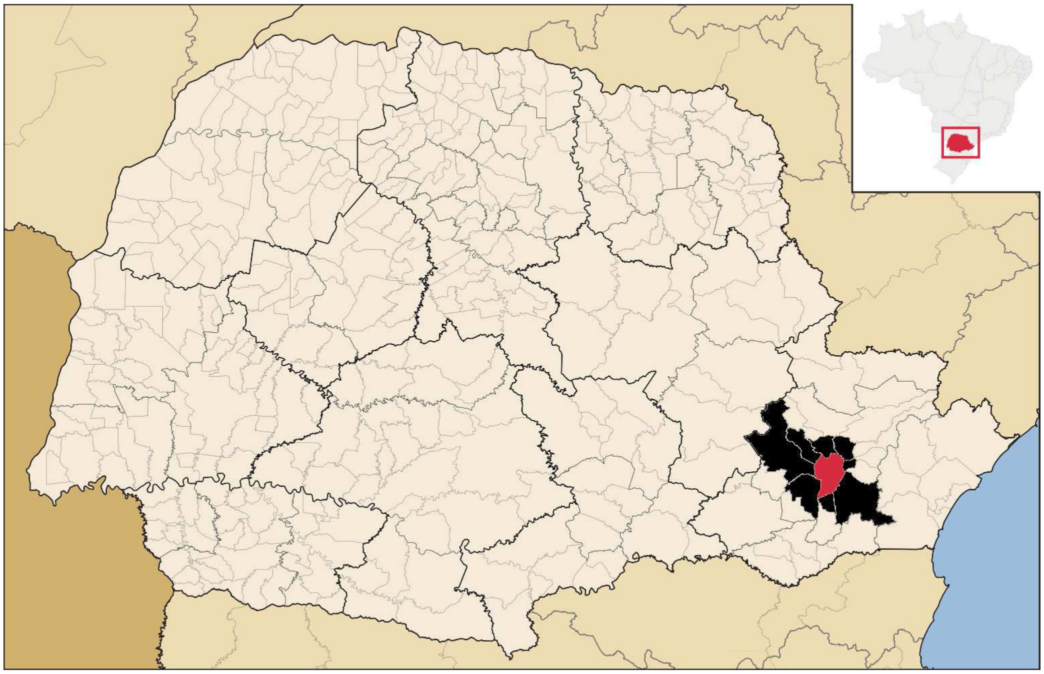


Figure 2: Task 2: Curitiba city and its neighbors.

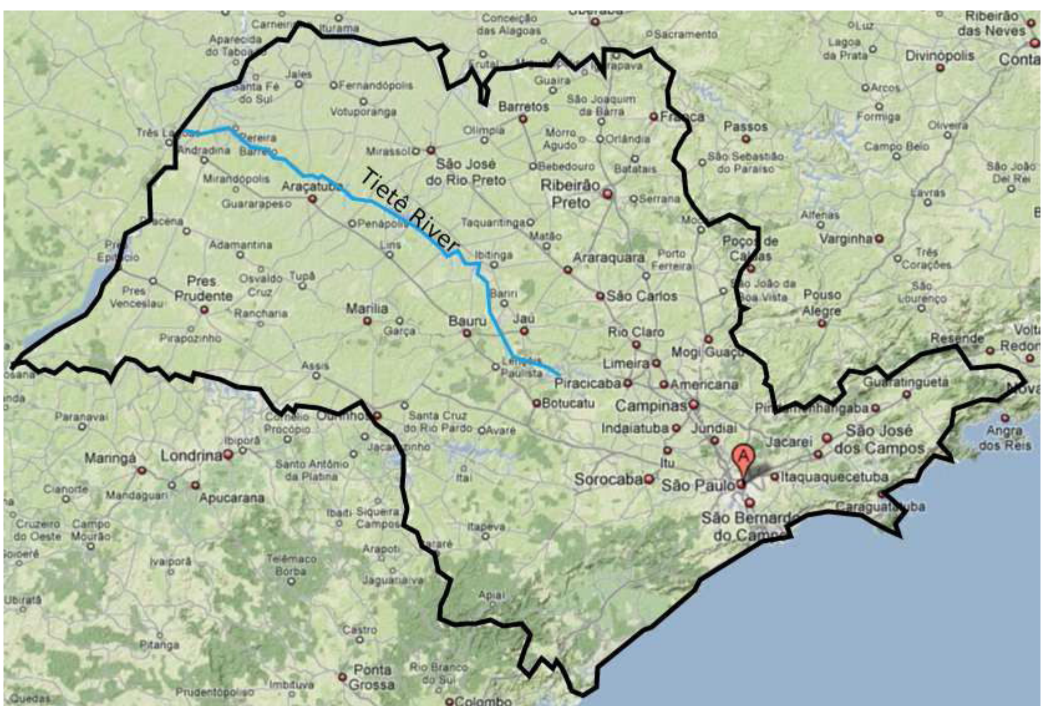


Figure 3: Task 3: São Paulo state and Tietê River.

Task context: You are going to Barcelona to attend a conference which will take place near a subway station and you would like to take the opportunity to know this city. It would be helpful to stay in a hotel near any subway station in order to make it easy to visit any piece of the city.

Task 5: Search for Web pages of cities of São Paulo State which are at the South of Campinas (see Figure 4).

Task context: You work in a digital inclusion project which will try out the latest generation communication technology connecting Campinas to Peruíbe, a city on the shore at south of Campinas. You wonder if more cities could also benefit from this same link. As a result you have the idea of contacting the city halls at south of Campinas which the transmission cable would potentially pass by in its way from Campinas to Peruíbe.

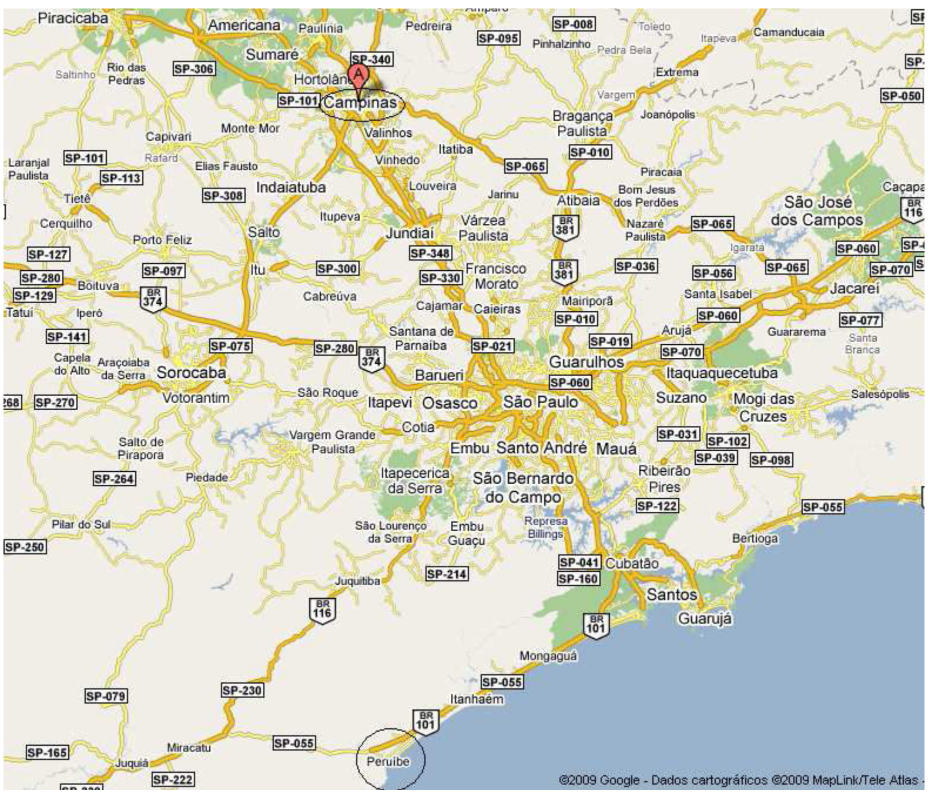


Figure 4: Task 5: Campinas and Peruíbe are higlighted on the map.

The relationship mentioned by task 1 actually leads to the objects marked in red in Figure 1. Figure 2 shows the geographic objects defined in task 2. In Figure 3, the boundaries of São Paulo State has black contours. It can be observed that Tietê river crosses the state from northern to southeast. Finally, the positions of Campinas and Peruíbe are highlighted on the map in Figure 4.

Opening questionnaire: At first, users answered an opening questionnaire about their familiarity with computers, use frequency, and main purpose of using computers; their knowledge on using search tools and use frequency; finally, they gave information about their formal education, as well as their English skills.

Main questionnaire: For each task, it was asked to the user to take note of and assess: (a) the steps used (or tried) to perform the task (it was asked to the user to describe how they attempted to perform it); (b) the execution time in minutes; (c) the satisfaction with the execution time; (d) the list of interesting links retrieved; (e) the satisfaction with the retrieved results; (f) the difficulty to perform the task; (g) in case of not being able to perform the task, it was asked to present the reasons for that.

Closing questionnaire: It was given the chance to the users to list examples of sites or Web tools they have liked, found interesting, or was useful to what they have done. Users were also able to leave their comments, criticisms, or suggestions about the experiment they had just participated.

3 Results and Analysis

Thirty four users accepted to join the experiment and their profiles were drawn from their answers to the opening questionnaire. It was a considerable achievement to get so many users to participate in this experiment just by considering an invitation letter. On median, users took about a half hour to answer the questionnaires.

3.1 Users' profile

Most of the users have *a lot* of computer skills (79%) and use computers *a lot* (85%), as shown in Figure 5, which also pictures that a great number of users know and use intensively Web search engines. In general (73%), the computer is used for work purpose (Figure 6). Additionally, more than a half of them have master/PhD degree or specialization studies as shown in Figure 7.

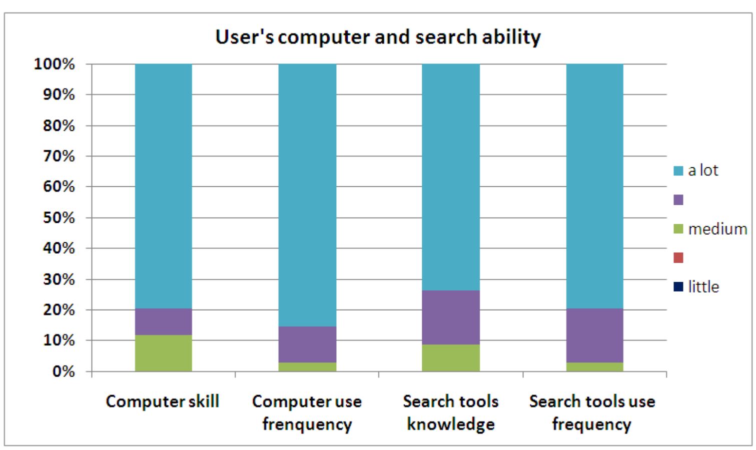


Figure 5: User's skill for computer and search tools and their use frequency.

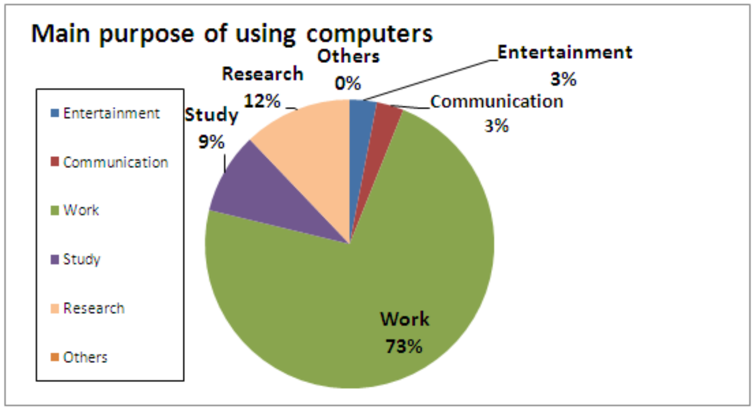


Figure 6: User's main purpose of using computer.

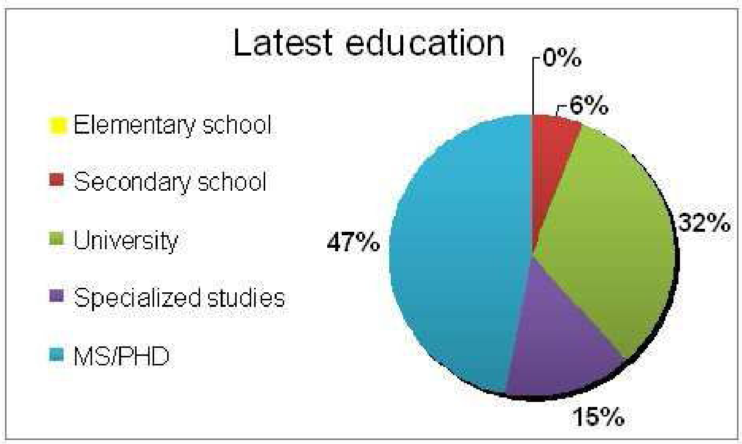


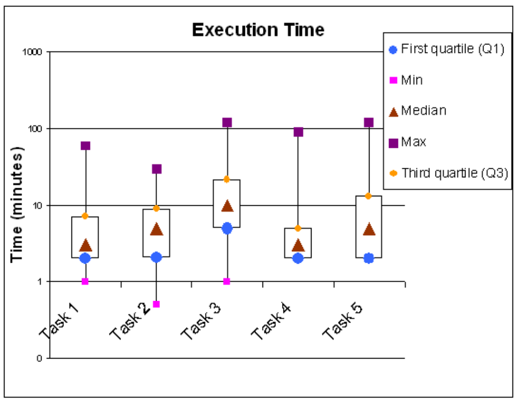
Figure 7: User's formal education.

As a conclusion, the users who participated in this experiment could be considered advanced users who use intensely search engines, are very familiar to computer as well as are highly educated. Thus, it seems these users have broad knowledge and are familiar with tools for information search on the Web.

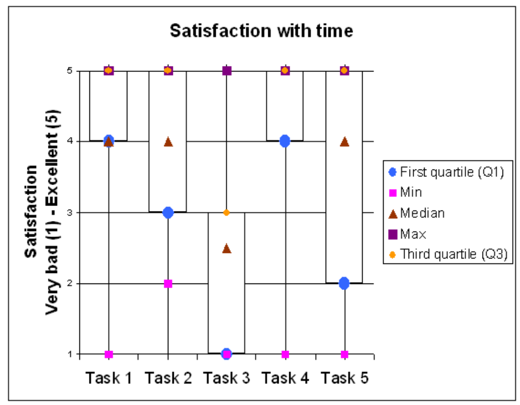
3.2 Main questionnaire results

The median time to execute the tasks ranged from 3 to 5 minutes, except for tasks 3, whose median time was 10 minutes (Figure 8(a)), though for tasks 3 and 5 the time far exceeded the median (120). Coincidentally task 3 was pointed out as the most difficult, followed by task 5 (Figure 8(d)). Users also evaluated task 3 as the worst in terms of time and result satisfactions as shown by Figure 8(b) and Figure 8(c). It was completed by only 22 out of 34 users as shown in Figure 9.

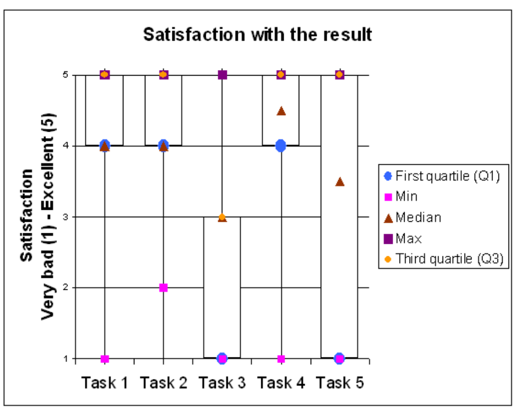
As it can be observed in Figure 8(b), to perform some tasks, around 3 to 5 minutes (tasks



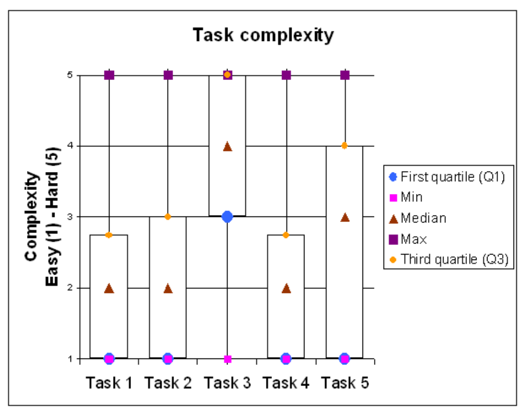
(a) Execution Time



(b) Satisfaction with the execution time



(c) Satisfaction with the results



(d) Complexity

Figure 8: Users' tasks rates

1, 2, and 5) was considered acceptable, since the users rated them 4 out 5, on median, for their satisfaction with the time. Note also that for task 3, users were not satisfied with the time spent. In fact, task 3 was considered time-consuming and complex (Figure 8(d)). That might be the reason for many users having not finished this task.

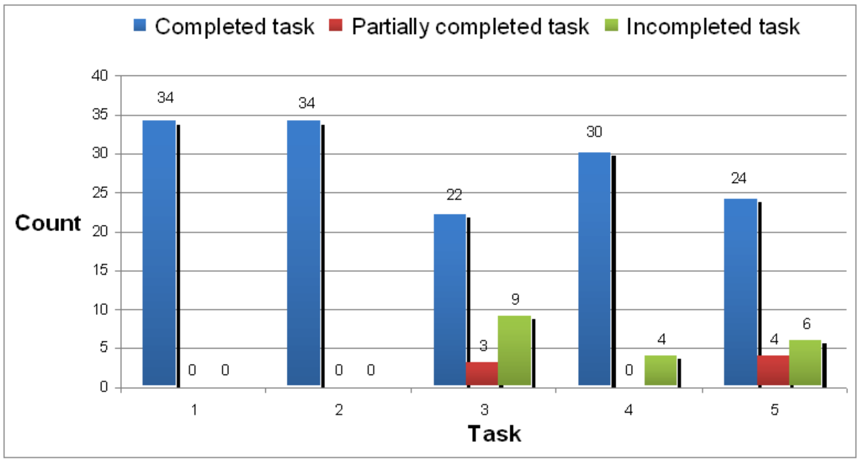


Figure 9: Tasks Completeness.

Additionally, the results show that Web tools can cope with some specific objects and relationship about which terms can be found on Web pages. For example, for task 4, the word "near", representing the geographical relationship of proximity, and objects "hotel" and city ("Barcelona") are relatively frequent on Web pages. Specifically for this domain, hotel pages usually have the term "near" mainly for advertising purposes (e.g., pages of hotel usually inform if the hotel is *near* a subway station or landmarks of a city). That would explain why it was quite straightforward to find relevant results when users just sent the query "Barcelona hotel near subway station" to a search tool.

Moreover, for geographic objects like cities or some political divisions (state, country), users use tools like Google Maps to retrieve the location of objects on maps that would help to infer spatial relations and possibly narrow down the search results.

On the other hand, for a more specific domain, like the scenario presented in task 3, associated to a quite unusual object (*river*) and geographic relationship (*cross*), it was difficult to find the target pages on the Web. It might be explained by the lack of keywords of this domain embedded in Web pages .

The difficult to execute task 5 relies on the relationship represented by word "south" as well as by the ambiguity in how the user does interpret this spatial relationship. A great number of users reported that they used Google Maps to draw a route from Campinas to Peruíbe. By following this route, they expected to find out the cities of interest and then their Web pages. Some other users just tried to determine visually on Google Maps the cities located below Campinas.

Discussion Most of the tasks were accomplished with more than one step. We believe that users with more experience in searching on the Web knew that existing search tools would not answer properly a geographic query in a single step.

For sake of general discussion, let's take task 2 "Search for Web pages of cities in the neighborhood of Curitiba" as an example. The same principle descibed here applies to the other tasks.

It is not enough sending the city name "Curitiba" and a term which refers to the geographical relationship to a search tool, because it will just match the keywords with the page contents. For those who were got used to geographic queries, it is common to rewrite the query in a way that the search tools are able to retrieve relevant results.

As a rule, a more complex geographical query was broken down into two main steps. In the first step the geographic query is transformed in keyword searches by:

- using user's previous knowledge to associate a city to a region, or reaching the city page by using a URL naming pattern previously known (e.g., the URL of city's home page in Brazil is formed by *www.<city name>.<acronym>.gov.br*, where *<city name>* is the city name and <acronym> is the acronym of its state, so for Curitiba, city of Paraná (PR) state, its URL is *http://www.curitiba.pr.gov.br*);
- visiting previously user known Web pages (e.g., Wikipedia). From these Web pages, users could find the cities near by, or the distance between cities. In this case, users go first to the Wikipedia to find the city of interest and then create a list of cities;
- submitting other words to the search tool in order to return the list of cities;
- using a map service to localize a city used as reference, visually inspect, and manually create a list of cities that satisfy the target geographical relationship. For example, going to a mapping tool like Google earth or Google maps first, find the city of interest and visually find the neighbor city.

The second step consists in using the names of cities identified in the first step as keywords in subsequent searches.

3.3 Closing questionnaire results

Common examples of sites or Web tools cited as good and useful were Google Maps, Wikipedia, and Wikimapia. Users also mentioned that they like taking advantage of existing URL patterns (e.g., Web pages of São Paulo state cities follow the pattern *www.<city name>.sp.gov.br*). Some interesting comments from users that confirm the experiment results were:

- "I believe the user's previous knowledge about the domain and knowing another language would help on searches";
- "In general, it is needed to submit a lot of queries on Web before completing a task";

- "Combining search tools with tools like Google Maps are extremely interesting and powerful";

- "Although I think Google Search is very good, I believe that for some specific domains (e.g., flight tickets, hotels) there is more information available on the Web. Whenever the search became more specific, I think it is more complicated to get relevant information on the Web quickly, as is the case for task 3...";

- "It would be a good idea having a tool to draw a polygon on a map and perform searches to bring results belonging to that area... In addition, inserting information collaboratively on maps, as Wikimapia, and having tool to search for this information would also be useful";

- "It would be interesting if every URL had some patterns of formation (e.g., Brazilian cities pages)...";

- "I miss a tool to classify what we are looking for, e.g., to get phone number, address, I have to navigate throughout the entire site.";

- "It lacks more integration between Web tools that combine geographic information and other kinds of information...";

- "The experiment showed that there are more search dimensions that are not well solved with existing search tools.";

- "I would like a search tool which lets me build a query with logic operators...(with SQL power)".

4 Architecture proposal

The experiment described above confirmed our hypotheses that the processing of queries involving geographical relationships between geographical objects is complex, does not return enough relevant results, and is time-consuming. Based on the experiment results, we propose an architecture to extend current search tools aiming to integrate existing tools and to improve the geographical search experience. The proposed architecture is structured in three layers as illustrated in Figure 10.

In the presentation layer there is a human-computer interface on which a user can define and refine a query, visualize retrieved results. This layer is implemented by using external API to enhance the exhibition of information retrieved from the Web, for example, Google Maps API [13]. These APIs are provided externally by other sites to help developers to incorporate external features to their own site.

In the processing layer, there are the modules responsible for: *geo-parsing* of the terms used on the query; geo-coding query terms; query expansion; handling relevance feedback; and ranking Web pages based on their relevance. The search engine manager forwards a query to many Web searches (meta searches), combines the results, and sends those results ranked by their relevance to the presentation layer.

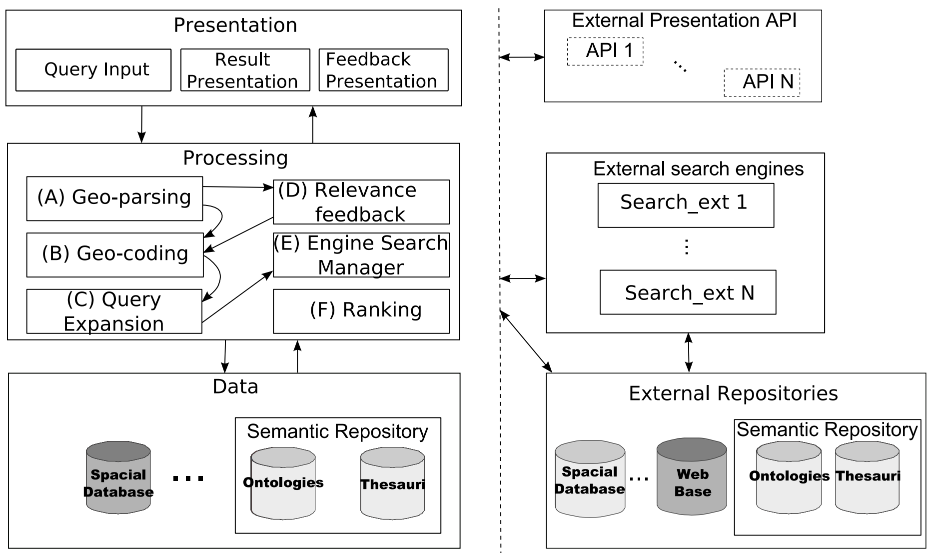


Figure 10: Architecture for geographical information retrieval on Web.

The data layer is composed of local repositories and those distributed by Web. These repositories contain data, ontologies, and thesauri to disambiguate terms and expand queries. The remote repositories include other ontologies and thesauri.

4.1 Scenario of use

A typical scenario of query would consist of the following steps: the user specifies a query; the system recognizes and disambiguates (geo-parse) terms referring to geographical objects in the query, for example, name of homonym places or names that refer to more than one object; the system might ask for user to filter out term senses and/or indicate the sense for terms used in the query. By means of the human-computer interface, the user indicates which sense should be considered. Next, the system geo-codes the referencing elements of the geographic query and expands the query which is then sent to the search engine manager. Before presenting the returned results, they are first combined and ranked by relevance. Finally, the user might want to filter out even more the presented results by indicating their relevance (user's feedback).

4.2 Prototype

Part of the proposed architecture (Figure 10) was implemented as a prototype. The implemented modules were Query input, Result presentation, Geo-coding, Query expansion, spatial database and communication functions which are in charge of using external presentation APIs.

Searches involving geographical relationship where implemented by sending queries first to a spatial database. The geographical relationship implemented was *touches* (neighborhood), *distance*, *crosses*, *inside*, *outside*, *intersects*, and *contains*. The spatial database was loaded with vectorial data downloaded from the Brazilian Institute of Geography and Statistic (IBGE) [16] about Brazilian cities and state boundaries, rivers, federal highways, and railways.

The query input is structured (Figure 11) guided by pre-defined fields. At first selection box, users indicate the kind of information (e.g., city's Web pages) they are interested in, the object type to which the information are related (e.g., city) – let call this object as aim-object. Next, it is selected the geographic relationship (e.g., distance) which those aim-objects are suppose to have with the object of reference (reference-object). For this, it is also specified what type is the spatial object, as well as it is named.

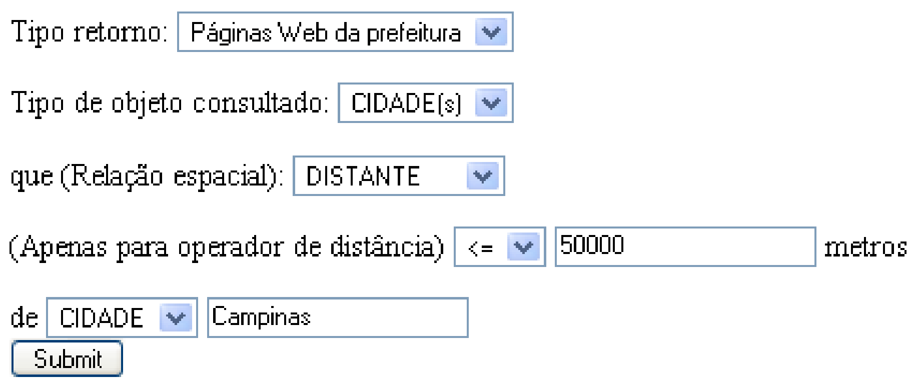


Figure 11: Interface to specify query with geographic relationship betweens geographic objects.

Once the reference-object is characterized, the geographical relationship is processed in the spatial database in order to retrieve the list of aim-objects. This list will help to expand the input query and the result of this expansion are sent to the Web search engines (in the current implementation, only the Google search engine is being used). The result is shown as a Web page which aggregates the list of returned results and shows the spatial localization of the aim-objects on a map (Figure 12). As described, the user is able to retrieve the information in just one step.

Let take Task 2 as an example to illustrate the implemented process. In this case, the target (searched) information is city's Web pages; the aim-objects include cities; the geographic relationship is *neighborhood*; and the reference-object is an geographic object named Curitiba. First, Curitiba is identified as the reference-object and is geo-coded. Next, a query is sent to the database aiming to retrieve a list of cities (aim-objects) which *touch* Curitiba. After that, the original query is expanded in such way that the spatial relationship query is withdrawn and is replaced by the name of cities in the list. Finally, the expanded query is sent to the search engines. After receiving the results back, each aim-object is highlighted on a map as well as the links to these target (cities') web pages are listed. The same process works for Task 3. In this case, the searched information is concessionaries pages, the aim-objects is highways, the geographic relationship is *cross*, the reference-object is river and its name is Tiête.

Figure 12: Result for que query "Which are the city's page of cities near (up to 50 km) city of Campinas?".

This prototype was implemented using Javascript and Python [28] under application Web server Django [8]. The search engine was provided by Google AJAX Search API [11] and the presentation of aim-objects on map was provided by Google Maps API [12]. The spatial data repository adopted was PostgreSQL [27] with spacial extension PostGIS [29], which was loaded with vectorial data from IBGE site [16].

5 Related work

The area of *Geographic Information Retrieval* (GIR) deals with new challenges derived from the management of geographic data in the traditional area of *Information Retrieval* (IR). The GIR area can be seen as an extension of IR area with the addition of geographic objects's data and their relationships. It combines research on information retrieval and geographic information systems, involving problems that vary from storage issues to friendly interfaces [30, 20, 19, 22].

Finding documents related to geographic entities requires identifying [33] and disambiguating geo-references semantically similar or imprecise (geo-parsing) [25, 10, 4, 24, 21, 34], dealing with vagueness of the language to define location [31], defining the location on Earth of documents or terms (geo-coding) [3, 5, 2], providing spatial browsing, supporting efficient query processing [6], constructing and indexing a database of places [26, 18], as well as ranking the results according to their relevance [35, 23].

Due to great importance of geographic information in users' life there are even some works trying to understand and characterize users needs for this kind of information [15].

Even though a lot of research has been conducted on GIR, there is still lack of support in Web searches that are related to geographic objects and geographic relationship.

Geographic (spatial) relationships are very important in several applications such as Agriculture, Biodiversity [32], Infrastructure, Health, E-government, and Risk Management. There are several spatial databases that support the specification and execution of this kind of query. Commercial and non-commercial DBMS, like Oracle and PostgreSQL, have been extended to manage spatial objects and to cope with some of their relationships by means of operators. Spatial queries are classified according to the kind of operator used to process them. A common classification distinguishes among topological, metric and directional queries [14].

Few researches have been conducted to support the specification and execution of Web queries involving spatial relationships. An interesting work is the one focused on modeling the geographical relationship between places to support geographic information retrieval using geo-ontology [1]. The main objective is to capture spatial relationships of adjacency and orientation between geo-objects.

Some initiatives on processing geographic relationships on GIR systems are:

SPIRIT – *Spatially-Aware Information Retrieval on the Internet* project, whose aim is to develop a spatially-Aware search engine. The spatial relationships supported initially are "inside", "outside" and "near", as well as cardinal direction and proximity relationships (within a specified distance) [18].

ADEPT – Alexandria Digital Earth ProtoType came from Alexandria Digital Library (ADL). It is a distributed digital library comprising of a collection of geo-referenced material that could be searched in [17, 9]. It does not perform Web search, but the spatial operators supported in searching the distributed digital library are: "contains item are", "overlaps", "encompasses/contained by", and "exclude/outside".

Google Maps provide a map and use Local Search to find places by address and latitude and longitude coordinates. This queries can also be combined with keyword searches. From a found point of interest (POI) on a map, it is possible to search for another POI "in the area" defined by the first one. However, it does not encompass the whole Web pages indexed by Google Search.

6 Conclusions

This work has described ongoing research to evaluate the limitations of existing tools to support geographic information retrieval on the Web. We carried out an experiment to assess how well queries involving geographic relationships can be specified and executed using existing Web tools. Our evaluation considered three different aspects: the relevance of the results and the time and complexity of the steps necessary to obtain these results. More than thirty Web users performed five tasks, being the users' profile considered advanced.

Experiment results demonstrate that existing Web tools are not enough integrated to meet user needs regarding the specification and execution of queries involving spatial relationships. Note that is true even for users with advanced profile, like those who performed this experiment.

In general, users are not fully satisfied with the results and with the time spent to perform this kind of queries. Furthermore, these complex queries usually require switching several times across different Web tools. Exceptions were identified when terms associated to geographic objects and/or spatial relationships are explicitly found on pages. In this case, the keywords-matching strategy, commonly used by Web search tools yields good results.

Another contribution of this work relies on the proposal of an architecture to process geo queries on the Web. As future work we plan to refine and fully implement this architecture aiming to integrate current Web tools so that they can support the execution of queries involving geographic relationships.

7 Acknowledgments

Ricardo Torres is supported by FAPESP, CNPq, and CAPES. Lin Tzy Li is supported by CPqD. The researches would like to thank the users who participated and contributed to this experiment.

References

[1] A. Abdelmoty and B. El-Geresy. Modeling spatial relations between places to support geographic information retrieval. In *Knowledge-Based Intelligent Information and Engineering Systems*, volume 5178/2008, pages 689–694, Zagreb, Croatia, 2008. Springer.

[2] D. Ahlers and S. Boll. Retrieving address-based locations from the web. In *Proceeding of the 2nd international workshop on Geographic information retrieval*, pages 27–34, Napa Valley, California, USA, 2008. ACM.

[3] K. A. V. Borges, A. H. F. Laender, C. B. Medeiros, and J. C. A. Davis. Discovering geographic locations in web pages using urban addresses. In *Proceedings of the 4th ACM workshop on Geographical information retrieval*, pages 31–36, Lisbon, Portugal, 2007.

[4] D. Buscaldi and P. Rosso. A conceptual density-based approach for the disambiguation of toponyms. *International Journal of Geographical Information Science*, 22(3):301, 2008.

[5] C. E. C. Campelo and C. de Souza Baptista. Geographic scope modeling for web documents. In *Proceeding of the 2nd international workshop on Geographic information retrieval*, pages 11–18, Napa Valley, California, USA, 2008. ACM.

[6] Y.-Y. Chen, T. Suel, and A. Markowetz. Efficient query processing in geographic web search engines. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 277–288, Chicago, IL, USA, 2006.

[7] E. Clementini, P. D. Felice, and P. van Oosterom. A small set of formal topological relationships suitable for end-user interaction. In *Symposium on Advances in Spatial Databases*, pages 277–295. Springer-Verlag, 1993.

[8] Django Software Foundation. Django: The web framework for perfectionists with deadlines. http://www.djangoproject.com/.

[9] M. Freeston. The alexandria digital library and the alexandria digital earth prototype. In *JCDL '04: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, pages 410–410, New York, NY, USA, 2004. ACM.

[10] G. Fu, C. B. Jones, and A. I. Abdelmoty. Ontology-based spatial query expansion in information retrieval. In *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE*, Lecture Notes in Computer Science, pages 1466–1482. Springer, 2005.

[11] Google. Google ajax search api - google code. http://code.google.com/apis/ajaxsearch/.

[12] Google. Google maps api reference - google maps api - google code. http://code.google.com/apis/maps/documentation/.

[13] Google. Map basics - google maps api - google code. http://code.google.com/apis/maps/documentation/.

[14] R. H. Guting. An Introduction to Spatial Database Systems. *The International Journal on Very Large Data Bases*, 4(3):357–399, October 1994.

[15] A. Henrich and V. Luedecke. Characteristics of geographic information needs. In *Proceedings of the 4th ACM workshop on Geographical information retrieval*, pages 1–6, Lisbon, Portugal, 2007.

[16] IBGE. Mapas interativos - ibge. http://www.ibge.gov.br/mapas/.

[17] G. Janée and J. Frew. The ADEPT digital library architecture. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 342–350, Portland, Oregon, USA, 2002. ACM.

[18] C. B. Jones, A. I. Abdelmoty, D. Finch, G. Fu, and S. Vaid. The SPIRIT spatial search engine: Architecture, ontologies and spatial indexing. In *Geographic Information Science*, Lecture Notes in Computer Science, pages 125–139. Springer, 2004.

[19] C. B. Jones and R. S. Purves. Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3):219, 2008.

[20] T. Kobayashi, K. Misue, B. Shizuki, and J. Tanaka. Information gathering support interface by the overview presentation of web search results. In *APVis '06: Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation*, pages 103–108, Darlinghurst, Australia, 2006. Australian Computer Society, Inc.

[21] J. Leveling and S. Hartrumpf. On metonymy recognition for geographic information retrieval. *International Journal of Geographical Information Science*, 22(3):289, 2008.

[22] J. Machado, G. Pedrosa, N. Freire, B. Martins, and H. Manguinhas. User interface for a Geo-Temporal search service using DIGMAP components. In *Research and Advanced Technology for Digital Libraries*, volume 5714/2009 of *LNCS*, pages 483–486, Corfu, Greece, 2009. Springer.

[23] B. Martins, M. J. Silva, and L. Andrade. Indexing and ranking in geo-IR systems. In *GIR '05: Proceedings of the 2005 workshop on Geographic information retrieval*, pages 31–34, New York, NY, USA, 2005. ACM.

[24] S. Overell and S. RÃ¼ger. Using co-occurrence models for placename disambiguation. *International Journal of Geographical Information Science*, 22(3):265, 2008.

[25] R. C. Pasley, P. D. Clough, and M. Sanderson. Geo-tagging for imprecise regions of different sizes. In *Proceedings of the 4th ACM workshop on Geographical information retrieval*, pages 77–82, Lisbon, Portugal, 2007.

[26] A. Popescu, G. Grefenstette, and P. A. M. ellic. Gazetiki: automatic creation of a geographical gazetteer. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, pages 85–93, Pittsburgh PA, PA, USA, 2008.

[27] PostgreSQL. PostgreSQL: The world's most advanced open source database. http://www.postgresql.org/.

[28] Python Software Foundation. Python programming language – official website. http://www.python.org/.

[29] Refractions Research. PostGIS: Home. http://postgis.refractions.net/.

[30] P. Rigaux, M. O. Scholl, A. Voisard, and J. Gray. *Spatial Databases: With Application to GIS*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, California, CA (EUA), 2002.

[31] S. Schockaert, M. D. Cock, and E. E. Kerre. Location approximation for local search services using natural language hints. *International Journal of Geographical Information Science*, 22(3):315, 2008.

[32] R. S. Torres, C. B. Medeiros, M. A. Goncalves, and E. A. Fox. A Digital Library Framework for Biodiversity Information Systems. *International Journal on Digital Libraries*, 6(1):3–17, February 2006.

[33] F. A. Twaroch, P. D. Smart, and C. B. Jones. Mining the web to detect place names. In *Proceeding of the 2nd international workshop on Geographic information retrieval*, pages 43–44, Napa Valley, California, USA, 2008. ACM.

[34] F. A. Twaroch, P. D. Smart, and C. B. Jones. Mining the web to detect place names. In *GIR '08: Proceeding of the 2nd international workshop on Geographic information retrieval*, pages 43–44, New York, NY, USA, 2008. ACM.

[35] B. Yu and G. Cai. A query-aware document ranking method for geographic information retrieval. In *GIR '07: Proceedings of the 4th ACM workshop on Geographical information retrieval*, pages 49–54, New York, NY, USA, 2007. ACM.