

Geographic Information Retrieval and the World Wide Web: A Match Made in Electronic Space

David Johnson, MLS

and Myke Gluck, Ph.D.

Florida State University

School of Library and Information Studies

244 Shores Building

Tallahassee, FL 32306-2048

Email: djohnso@mailers.fsu.edu Or mgluck@lis.fsu.edu

ABSTRACT

This article looks at the access to geographic information through a review of information science theory and its application to the WWW. The two most common retrieval systems are information and data retrieval. A retrieval system has seven elements: retrieval models, indexing, match and retrieval, relevance, order, query languages and query specification. The goal of information retrieval is to match the user's needs to the information that is in the system. Retrieval of geographic information is a combination of both information and data retrieval. Aids to effective retrieval of geographic information are: query languages that employ icons and natural language, automatic indexing of geographic information, and standardization of geographic information. One area that has seen an explosion of geographic information retrieval systems (GIR's) is the World Wide Web (WWW). The final section of this article discusses how seven WWW GIR's solve the the problem of matching the user's information needs to the information in the system.

INTRODUCTION¹

Information retrieval (IR) systems attempt to satisfy requests for either general information by providing documents on a topic, or precise data from a database. Geographic information requests are often a blend of the two. Information can come in a variety of formats but the critical measure of success is the relevancy of the retrieved documents to the information needed. The solution to the traditional information

¹ This article deals with information science issues pertaining to geographic information, such as the retrieval of information and ability to effectively query a retrieval system for information. This is not a technical article dealing with design or interface issues of specific geographic information systems. Pertinent articles not discussed are included after the reference section.

retrieval problem is a system's best match between a user's information need and the system-held documents that resolve the information need. One very problematic issue for the retrieval of geographic information is query formulation. Retrieval systems currently have limited support to operationalize a user's natural language or other geospatial queries as well as limited support for users to express their information needs to the system. Fortunately there are alternative query languages beyond that of basic boolean operators of "anding" and "oring" keywords.

One alternative is "iconic query language" whereby a user creates a geographic query by using icons. An icon is a computer generated representation of a physical object; for example, a line representing a road or stream. Geographic information deals with physical objects that are in some cases hard to express with words. The ability to graphically create the query may facilitate more effective retrieval. Traditional query languages deal only with the precise (Cats AND Dogs) and cannot handle the vagueness of concepts such as "extremely low humidity" or "recent information." Users do not always think in precise terms, which can also make it difficult to retrieve information. By using a fuzzy query language, the concept can be converted into a number sequence that the system can use. Similarly, automatic indexing of georeferenced information can also help a user to retrieve geographic information.

The introduction of the World Wide Web (WWW) and clickable image maps have simplified the users' search and retrieval for geographic information. The relative infancy of the WWW and online geographic information systems (GIS) currently limit most GIR's to queries for specific factual data, further constrained from choices only from menus. Relatively few WWW GIR systems allow users to edit and save geographic information online. We examined seven WWW GIS systems to discover how "the match" between a user's information need and the system response for information was accomplished. We present our results below.

INFORMATION RETRIEVAL

This is the age of information. Year by year the amount of data and information increases exponentially with geospatial data accounting for a large proportion of this increase. Too much information can cause an information overload driving the need to weed out the irrelevant information in order to retrieve the relevant information. In information science, the discipline that deals with this subject is information retrieval or IR. IR deals with “the process of searching some collection of documents, using the term documents in its widest sense in order to identify those documents which deal with a particular subject” (Lancaster 1979, 11). Most IR systems and research attempt to resolve the classic IR matching problem mentioned earlier: The systems try to efficiently and effectively “match” a user's information need to the document or documents that can fulfill that need. Typical commercial IR systems use a Boolean-based query mechanism and an inverted file of terms and documents (e.g., DIALOG or LEXIS/NEXIS). There are other types of IR systems such as Salton’s SMART system² based on vector space and Belkin’s ASK (“Anomalous States of Knowledge”)³ system which instead of trying to get the perfect “match” tries to get the user close enough to the right documents, similar to browsing in a library stack.

According to Ray R. Larson (1996), six basic elements make up a retrieval system. All six retrieval elements are pertinent to the retrieval of geospatial data and information and will be addressed in the Geographic Information and Information Retrieval section of this paper. The six elements are: (1) retrieval models, (2) indexing, (3) how items are matched and retrieved, (4) relevance, (5) order of results, and (6) query

² Salton, G. & McGill, M.J. 1983. *Introduction to modern information retrieval*. New York: McGraw-Hill.

³ Belkin, N.J. & R.N. Oddy & H.M. Brooks 1982. ASK for Information Retrieval: Part I. Background and Theory. *Journal of Documentation* 38(2):

languages. Van Rijsbergen (1979) adds query specification as a seventh element. With these seven elements, one can clarify the differences between retrieving information, and retrieving just data; that is, the distinction between IR and data retrieval (DR).

Retrieval Models

The first element of the retrieval system is the mechanism or model the system uses. Some IR systems are based upon a probabilistic or approximate model that deals with “subjective issues”(Larson). This model finds information relevant to the user although the document may not satisfy all of the user’s requirements. However, DR (data retrieval), on the other hand, uses a deterministic model retrieving documents that contain the stated variables as required by the query. For example, DR is concerned with queries like seeking the number of widgets in the inventory at the Chicago plant. IR queries are concerned more with finding some information on widget construction or regional differences in processes concerning widgets.

Indexing

Indexing organizes the documents of a retrieval system to promote efficient and effective intellectual as well as physical access. In IR systems, indexing is “derived from the contents, “ (Larson) such as keyword extraction. IR can also index terms from a controlled vocabulary (Pao, 1989) which is a preselected list of words that can be used to describe any document in the collection. In DR indexing the item itself is the indexing unit. For example, an automobile part in a parts database uses the part number or name as the indexing unit or entry point into the index.

Match and Retrieval

When a retrieval system attempts to connect a user’s information need to the documents that may resolve the need, the system matches and then retrieves the needed documents or pointers to the documents (page numbers, report titles, etc.). IR and DR use two different matching methods. IR uses a “partial match, best match” (Van Rijsbergen 1979, 2) method where the idea is “to find those items which partially match

the request and then select from those a few of the best matching ones” (Van Rijsbergen 1979, 2). The IR system attempts to exhibit just those documents that can solve the information need. A retrieved document might not be the ideal, but it will be close enough (at least initially from the system’s point of view). DR, however, uses an “exact match” method. When a particular datum is sought, the DR system returns that datum, or if not contained in the system, returns nothing. DR systems are extremely precise but brittle and fail easily without signs of degrading.

Relevance and Ordering

Larson’s fourth and fifth elements pertain to the display of information. Each retrieval system has a “criteria” for the items retrieved. In the IR system, relevance is the reason that the items are retrieved. An item is considered relevant if it satisfies all the requirements of the user. Again using the widget example, a user wants information on the widget construction done in a Chicago plant. An item retrieved that deals with that exact topic will be considered extremely relevant, whereas a document generally on widget construction is relevant, but not as relevant as the first item. Items that are retrieved by an IR system are ranked in a particular order, the most relevant items at the top of the list and the least relevant at the end. Ordering of the items retrieved by an DR system, on the other hand, are arbitrary without preference to the degree of relevancy of the item.

Query Language

The last two elements are Larson’s query language element and Van Rijsbergen’s query specification element. Both elements deal with the query mechanism of the retrieval system. In an IR system, queries are more natural-language like with less structure than the query language of a DR (e.g., Structured Query Language or SQL). A natural language query can be as simple as entering one or two sentences into the system or as complex as a series of paragraphs. The query mechanisms for IR systems are intentionally incomplete (Van Rijsbergen 1979). The user may either lack the ability to fully articulate the query or (since information needs are dynamic) not anticipate the

full specifications, breadth, depth or precision of the information needed. Either condition may result in an incomplete formulation of the query. DR uses a query language such as SQL that is artificial and more structured, employing more restricted syntax and vocabulary (Van Rijsbergen 1979). The specifications of a DR query are a complete specification of the desired item.

GEOGRAPHIC INFORMATION AND INFORMATION RETRIEVAL

The retrieval of geographic information attempts to provide intellectual and physical access to geo-referenced information sources. Unfortunately, geographic information is not easily compartmentalized into just one retrieval system. Larson uses the six elements discussed above to investigate Geographic Information Retrieval's (GIR's) position on the "retrieval continuum." One end of the retrieval continuum is DR while on the opposite side is IR. GIR, coined by Ray R. Larson of U.C. Berkeley, is a combination of the information retrieval and data retrieval elements discussed in the previous section.

GIR's "retrieval model" is a combination of deterministic and probabilistic models. Retrieval of geographic information can range from searching out a single location, to finding all towns near a major highway spanning more than one dataset, to finding likely locations in disparate regions for further investigation. Indexing of geographic information also blends information and data retrieval. GIR can use elements from the item, such as "intellectual" or "inferential" indexing. GIR indexing may also be of the complete item, as in data retrieval, such as a map or data set. In matching and retrieval, GIR can use both, the "partial or best match" of IR and the exact match of DR. A "partial or best match" would be finding information on the towns near a major highway that match imprecise or value laden attributes (e.g., near, not too far from, etc.). GIR also uses the "exact" match method of data retrieval. For example, requesting data on carbon dioxide levels for a range of times, dates, and elevations at a particular place.

Standardization

One important means to facilitate the retrieval of geographic information is through standardization. The United States government has spent a vast amount of money compiling geographic data and standardization is seen as a way to cut spending and reduce data duplication. The Federal Geographic Data Committee (FGDC) has been working toward coordinating all federal “geographically referenced information” (Nerbert). In 1994, the FGDC was a partner in creating the National Spatial Data Infrastructure (NSDI) to help coordinate spatial data access. The NSDI will be the “Clearinghouse” for all the U.S. government geographic data. NSDI’s functions are:

... developing and implementing standards for the framework and thematic data; producing framework and thematic data; implementing standards for geospatial data documentation and transfer; establishing procedures to use electronic networks to search for, access, and use geospatial data; and cooperating in the development of state and regional councils and private sector agreements to accomplish these actions.⁴

The NSDI made freeWAIS, a Wide Area Information Server (Nerbert), the standard platform to be used by all governmental agencies for indexing and serving geographic data. Standardization of “geospatial” data will allow more effective search and retrieval of geographic information via the WWW or with GIR systems.

Experimental Query Languages

Geographic information needs can be difficult to express verbally, creating problems for both system and user when trying to form a query. The inability to accurately portray a query to a system will result in poor matches or the retrieval of incorrect data. Two innovative querying languages for GIR systems are iconic and fuzzy query languages. Both languages support improved query formulation and translation for system matching. The ability to express the information needs graphically can improve the matching capability of the system. Iconic query languages create

⁴ A more complete description of NSDI may be found at URL

graphical query icons. Lee and Chin (1995) created an iconic query language to simplify retrieval of geographic information from a spatial database. The icons used to build a query were lines, points, boxes, and asterisks. With a graphical user interface (GUI), the user can choose the icons to represent the geospatial information need. The GUI is linked to a database interface that is in turn linked to a relational database (INGRES). The database interface is used to translate the iconic query language into an expression the relational database can employ.

The authors provide an operational example to show the effectiveness of an iconic query language user interface. This system is still in the early stages of development. There are also suggestions for improvements to future user interfaces, such as: the need for a wider choice of icons to choose from and a better user verification system (Lee and Chin, 1995).

No IR system permits or is capable of full natural language interaction with users. Users must know the system commands allowed for effective query formulation. Often with the more traditional IR systems (DIALOG, LEXIS/NEXIS, etc.) the commands are difficult to use or understand. Systems would be more user friendly if a user could enter a sentence or two in their own words to explain the information needed.

Natural language is the ability for the system to consider the sentence as a whole rather than the interpretation of individual words. If one enters a query such as “I would like recent information on Bosnia,” what should be retrieved? The pertinent information would not be a document that has the words “recent” or “information,” but a document that has recently been written about Bosnia. The ability of a system to handle these nuances has potential for improving retrieval of geographic information. Fangju Wang has created a natural language user interface (Wang, 1994) that uses fuzzy query language. The fuzzy query language can handle modifiers such as “almost near” and “much higher than average.”

The fuzzy query language uses fuzzy formulae to handle the modifiers. The basic fuzzy formulae would look something like this: X is F, where F is the “membership

function,” such as “high.” Wang (1994) describes four classes of fuzzy formulae: (1) simple (temperature is low); (2) modification (temperature is very low), (3) comparison (“temperature is *slightly* higher (greater) than 10 0 C “ (Wang 1994, 149)) and (4) quantitative (“Find all the cities which are close to at least several camping sites,” (Wang 1994, 149)). This prototype system was tested with ARC/INFO using a fuzzy SQL. The system translates the fuzzy SQL queries into standard SQL, so the database can be queried.

The author tested his fuzzy query user interface on an existing GIS of Canadian tourist and weather information. The sample queries followed the SQL format of “select-from-where”(Wang 1994). These queries varied in the complexity of the “selection conditions” of the “where” part of the SQL form. Preliminary testing of the user interface by Wang has shown that fuzzy query language interface has the potential to improve access to geographic information.

Indexing Geographic Information

Indexing documents is a major part of any retrieval system. Indexing describes a documents content then formulates a surrogate for the document. Indexing in IR systems determines what are acceptable terms (Pao 1989) for search and retrieval of information. Geographic terminology can be used in geographic information systems as indexing terms. Two projects that are trying to index automatically “geo-referenced” information are: GIPSY (Geo-referenced Information Processing System (Woodruff and Plaunt, 1994)) and Mark Carlotto’s HyperMap system (Carlotto, 1995).

GIPSY is a geographic indexing system that assigns index entries to text documents. Each indexing entry is a polygon previously associated with the text term from base maps. A user can input a query (in the form of a paragraph) and GIPSY will assign the polygons from a base map that relate to those terms. All the individual polygons associated with the text terms are then overlaid on the map forming a stack. Regions with the highest stacks are seen as the most relevant regions to the user’s query. The query polygons are then matched to the stored document indexing to find

the documents that are most relevant to the user's query. Those with the highest relevance are returned to the user as an answer set.

To assign index polygons, GIPSY first parses words in the text of the query or the document then extracts all geographical terms. The terms pulled out by the parser are matched to geographical terms found in a thesaurus. The spatial regions from the retrieved data sets are processed and then given a weight. The four used criteria to determine relevance weighting are: (1) geographic terms that were extracted, (2) location and frequency of the terms, (3) "knowledge of the geographic objects in the database and their attributes," (Woodruff and Plaunt 1994, 649) and (4) "spatial reasoning about the geographic constructs occurring in phrases" (Woodruff and Plaunt 1994, 649). The final step is to stack the polygons. Here the polygons are mapped using the weighted information, the higher the polygon is stacked, the higher the relevance.

Mark Carlotto (1995) proposes the use of textual attributes and text processing for indexing and retrieving geographical information. In this system, text documents and queries that deal with geographic information are converted into binary vectors. Using various distance measures between the vectors (i.e., hamming distance), the system compares documents and queries. Because text documents that are less similar will have higher Hamming distances than documents that are more similar, distance can be used as a measure of relevance. Relational data can also be converted into "free text attributes" by concatenating the relational database field name with the field value (Carlotto, 1995).⁵

Carlotto used a prototype software program, HyperMap, to retrieve geographic information. HyperMap is a multimedia browser displaying geographic information such as a map, images, sound, and movies. HyperMap indexes the information by

⁵ An example of this process would be: "LandCoverTypeWater"

(Carlotto, 1995)

converting text files into ASCII format. A parser extracts relevant information (geographic coordinates, time, etc.) from the ASCII format and the full text is converted into a vector. When a query is entered into the system, the system ranks and selects objects based on a score (one, if all terms are in the text, and less than one otherwise). Spatial queries are treated similarly to text queries with objects closest to the query normalized to one.

WORLD WIDE WEB AND GEOGRAPHIC INFORMATION

With clickable maps, PERL scripts, and JAVA applets, the WWW offers the ability to do interactive geographic information retrieval. The end user can search and retrieve geographic information from home, school or the library. A search of WWW GIR systems leads to various WWW systems that search and retrieve geographic information. Seven WWW GIS systems will be discussed in the following section because of their innovative ways of addressing the IR matching problem for geospatial information needs.

The range of the geographic information that can be found on the WWW is limited. Most of this information is domain specific. For example, geographic information may only pertain to Australia or Canada or environmental information, and the user can only select information appearing on a menu. Some WWW geographic information retrieval systems do allow the user to enter search terms, but fewer allow the user to actually enter their own information into the WWW GIR system.

The WWW has given scientists the ability to share geographic information eliminating the extra time or money to collect redundant data. The Environmental Research Information Network (ERIN)⁶ is a WWW GIR dedicated to Australian environmental information. Scientists throughout Australia have access to the same

⁶ The URL is http://www.erin.gov.au/cgi-bin/spatial_interface.

information. Before the query is submitted to the system, the user must select an interface and a region for search. There are five interface choices: Mapsheet index, Minimum Bounding Rectangle (not yet operational), Biogeographic Regionalism, Major Catchment Region, and Named Regions. The mapsheet index interface was used to explore ERIN GIS. With the interface and regions selected, the user is given a PERL form to input the actual query. The user then chooses the data set themes and selects the dataset containment type (completely within the region, encompassing the region, or both). Users may also enter terms and use Boolean logic to broaden or narrow the query. The user is given the opportunity at the query screen to select which WAIS server to search and select the maximum number of items to be returned by the system's search.

After the user submits a less structured query, the system returns an SQL form of the query as well as all the query-matching data sets. The user can then selectively view maps of the data set coverages. Each map item is clickable providing metadata for the coverage (citation Information, online reference information, spatial domain keywords, data quality information, and so forth) along with all hypertext links to access additional text or geographic information. To view the actual data, the user may select the `online_linkage` URL.

The KINDS (Knowledge based Interface to National Data Sets) Project⁷ is a user-friendly interface to browse or search the Bartholomew digital mapset spatial data of the United Kingdom via the WWW. This project uses two different indexing techniques: thematic and geographic. The thematic index consists of themes such as danger zones, political boundaries, and scenic areas. The geographic index is based on the National Grid (map of the U.K. divided into fifty-five squares, each square with its own two letter code).

The user can query the KINDS Project using either a geographic map interface or

⁷ The URL is <http://cs6400.mcc.ac.uk/kinds>

a freetext search engine. The geographic interface is the National Grid map. The user simply browses the map image, then selects the cell of the Grid that is of interest (See Figure 1). Once the user selects an area of interest, a map with the selected area is highlighted and a list of thematic data about that region is retrieved.

The second interface to the system uses a freetext search request strategy. The user is given a form to fill in the text search terms using boolean logic (AND, OR) and truncation (*, asterisk), to help broaden or narrow the search strategy. A list of documents ranked by relevance score is retrieved with the more relevant documents appearing higher in the list. The user is given the Feature Class (such as political boundaries or road), the Area Code (a two-letter code for the National Grid), and the relevance score for each item retrieved. By clicking on the document, the user retrieves a map of the area containing the Feature Class and the specifications of that data layer.

The EWSE (European Wide Service Exchange)⁸ homepage employs an innovative method for discovering information about what EWSE resources are in a particular region. The user is given a menu of image maps of Europe to search for the area of interest. Once the user selects a particular map, the area is presented at higher resolution. If the area selected is not appropriate, a user may easily return to the map menu by selecting a button on the WWW browser. When the region of interest is found, the user can click a button and retrieve EWSE resources for the selected area.

A second query screen for EWSE then pops up and the user is prompted to select a search for jobs, workshops, meteorological stations, products, or all of the above as well as a display format for information retrieved. The user can choose to have the output formatted with the desired information sorted by date, name, or type. In addition, the user can click on the retrieved documents to discover what events are occurring in that region.

⁸ The URL is <http://ewse.ceo.org>

MapQuest⁹ is an interactive atlas produced by GeoSystems Global Corporation. This on-line atlas allows the user to “Edit” the map and has the ability to store searches, two features not seen in many WWW GIR systems. The user can search or browse through the interactive atlas through a user friendly menu system with five choices: Find, View, Options, Edit and File. .

Querying the MapQuest atlas is done by using the Find menu. The user is presented a query form in which to enter as much information as possible about a location or business of interest. The form consists of the name, address, city, state and zip code to be used to search for the business. After the query form is submitted, a map of the area is displayed with a pin marker at the business location. The retrieved map can be manipulated by zooming to streets and panning in cardinal directions of north, east, south, or west.

Enduser browsing in MapQuest is done with the View menu. The user is given a map of the continental U.S., and with the zoom and pan buttons can browse the U.S. to discover the desired information.

One of the most interesting features of MapQuest is its capability to edit the retrieved maps. Editing is permitted only after the user fills out a form and becomes a MapQuest member. The user can then customize the atlas data. In Edit mode, the user can add points of interest (See Figure 2) to the desired area of the map. The user clicks on a point of interest on the map and then fills out a form that describes that point. There are eighteen possible points of interest ranging from health care to entertainment. Once the form is filled out, the point is saved.

MapQuest’s last menu option is File. This feature allows users to store map searches for later use. This feature is unique for WWW GIS systems.

⁹ The URL is <http://www.mapquest.com>

The Australian Geodynamics Cooperative Research Centre (AGCRC)¹⁰ homepage uses GRASSLinks to connect users with geographic information via the WWW. GRASS (Geographic Resources Analysis Support System) is a powerful public domain GIS system created by the U.S. Army Corps of Engineers. GRASSLinks is a good example of a user-friendly GIS interface (See Figure 3). No command input is needed. Casual use of a GIS requires an easy-to-use query mechanism for the retrieval of geographic information. If the commands for the system are too complex, usability of the GIS will be low and only those who have the time to learn the system will use it.

The AGCRC WWW GIS, though domain specific, is a prime example of how easy it can be to search and retrieve geographic information via the WWW (See Figure 3). After starting GRASSLinks, the screen displays a menu for creating a map. To create the backdrop of the map, the user selects one of sixteen available raster maps. In the next menu box, the user selects from among twenty-five vector layers. Users also select the colors of the vector layers. A third menu allows the user to select the number of site maps to plot as overlay points. A choice of three colors to display on the site maps is also available. The last two choices on this screen are region selection and the size of the map to be retrieved and displayed.

The delay in time between query and retrieval of the map is minimal, even though map creation was done on the fly in real time. Once the map is retrieved, the user can zoom-in or out, and pan across the map. One inconvenience of these manipulations is the need for the system to completely regenerate the entire map.

¹⁰ The URL is

<http://www.ned.dem.csiro.au/AGCRC/4dgm/grasslinks/grasslinks.html>

The NAISMap (National Atlas Information Service)¹¹ is also a WWW interface to a GIS. This GIS deals with Canadian Spatial data and emphasizes geology, hydrology, the environment and biodiversity. NAISMap is a straightforward easy-to-use WWW GIS. Step by step query menus guide the user in creating the desired map. The first menu prompts the user to select as many map layers as desired from the thirty base layers.¹² NAISMap then orders the layers and gives the query back to the user for verification. Once the query is verified, it is submitted.

NAISmap's next screen displays all the selected layers permitting the user to change the colors of each data layer. When all the colors are selected, the user can again submit the query and wait for display of the retrieved map.

TIGER Map Service¹³ is a prototype browser that creates maps on-the-fly using TIGER/92 data. With the TIGER interface, the user can browse maps or input search terms to retrieve geographic information. Input of exact latitude and longitude provides a third option for search and retrieval of the desired geographic information at a particular location. To browse the TIGER GIS, you can change the preset values to the U.S., zoom in, and click on the desired geographic location. Once the user is close to the

¹¹ The URL is

<http://ellesmere.ccm.emr.ca/naismap/naismap.html>

¹² Selection depends upon the system running the WWW browser. For example when using Netscape 2.0 running under a Linux operating system we can select as many layers as needed. When the same GIR system was queried using Netscape on a DOS machine we can only select one layer at a time.

¹³ The URL is <http://Tiger.census.gov/cgi-bin/mapbrowse-tbl>

location, the four pan and zoom buttons can be used to find the exact location.

A text search from the U.S. Gazetteer is done by inputting the name of the city, the state, and optionally, the zipcode for the desired geographic location. The U.S. Gazetteer is a searchable index that retrieves both 1990 Census data and a hypertext link to a TIGER map. Once the location has been found by either browsing, searching a term, or latitude / longitude, a map of the area is retrieved . With a map displayed, the user can zoom in and out and select from pop-up menus forcing layers to be displayed or hidden. Another menu on the screen gives the user the ability to display selected levels and themes of census data. A legend underneath the map explains the meaning of the different colors or lines, and includes the map scale.

CONCLUSION

The ability to access and retrieve relevant information, whether in textual or geographic form, is important in today's information age. By creating retrieval systems that emphasize the IR elements discussed earlier (such as relevance, ranking, and natural language querying) systems will become more effective and user friendly. These techniques can be applied to systems that deal with geographic information.

Indexing of geographic information will create more efficient systems and better access. Data collected can be automatically indexed using systems like GIPSY or the prototype HyperMap. Terms can then be associated more easily with geographic locations. One possible future scenario would be to have access to all public geographic data from one GIR. The data, of course, could be distributed across many servers and would be accessed through a single user friendly interface built using the IR elements.

Another way to make the access to geographic information easier is through experimental query languages. Often, explaining geographic concepts in nature is difficult (Remember the last time you got directions and soon became lost in an unfamiliar city?) Therefore, why should users be limited to creating a query using text and boolean operators? Access can also be enhanced with natural query languages. The fuzzy query language user interface has the ability not only to improve access to

geographic information but also traditional IR systems. A user could ask for documents that are close to a particular call number or that have been published recently. A fuzzy query can make natural querying easy and effective for the user and the system.

With the growing acceptance of the WWW, the public will want access to more geographic information. Therefore, the user will need effective access and retrieval of geographic information. The WWW GIR systems discussed in this paper have begun to explore components to help in the retrieval of geographic information. Most of the WWW GIR's discussed permit users to browse the map for information or to present a query screen to enter search terms. These two methods are both effective ways to access information. Menu-driven GIR systems are a mixed blessing. Menu-driven systems are a success because they are easy to learn and use; all one does is select from the numerous menus and then have the system create the map. But, menus represent static or canned queries that do not allow unanticipated or flexible query formulation, since the user can only select from the menu items.

The seven WWW GIS's discussed are innovative methods for access and retrieval of geographic and textual information. MapQuest offers the ability to edit and save searches, something few search engines offer for free. Vendors such as DIALOG and LEXIS/NEXIS offer such amenities for text based searches but at a cost. The ability to customize the geographic information can be highly beneficial for a user. The Tiger Map Service does offer users the ability to modify images by placing a colored stick pin on the map; however, this is quite limited compared to MapQuest's customability. EWSE's ability to access textual along with pure geographic information is quite effective since a geographic browser should permit users to access all sorts of information.

GIR's are improving on the WWW, but better systems are still possible. From an IR standpoint, not all of the six elements were used by any of the WWW GIR's. Natural language queries are still not possible and most systems do not retrieve information ranked by relevance (only one of the seven retrieved textual information by relevance).

In the years to come, as more geographic information is put up on the WWW, users will find more pertinent information and it will be essential to span the domains of knowledge. Geographic information can also take time to download especially if in audio or video formats. GIR needs efficient and effective transfer and display of large datasets. Improvements as mentioned above will further assist users of geographic information to meet their information needs in the years to come.

REFERENCES

- Belkin, N.J. & R.N. Oddy & H.M. Brooks 1982. ASK for Information Retrieval: Part I. Background and Theory. *Journal of Documentaion* 38(2): 61-71.
- Carlotto, M.J. 1995. Text attributes and processing techniques in geographical information systems. *International Journal of Geographical Information Systems* 9(6): 621-635.
- Lancaster, F.W. 1979. *Information Retrieval Systems*. New York: John Wiley & Sons.
- Larson, R.R. Geographic Information Retrieval and Spatial Browsing. http://sherlock.berkeley.edu/geo_ir/PART1.html
- Lee, Y.C. & Chin, F.L. 1995. An iconic query language for topological relationships in GIS. *International Journal of Geographical Information Systems* 9(1): 25-46.
- Nerbert, D.D. Serving digital map information through the World-Wide Web and Wide-Area Information Server technology. <http://h2o.er.usgs.gov/WWW.paper.html>
- Pao, M.L. 1989. *Concepts of Information Retrieval*. Englewood, Colorado: Libraries Unlimited, Inc.
- Salton, G. & McGill, M.J. 1983. *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Wang, F. 1994. Towards a natural language user interface: an approach of fuzzy query. *International Journal of Geographical Information Systems* 8(2): 143-162.
- Woodruff, A.G. & Plaunt, C. 1994. GIPSY: Automated geographic indexing of text documents. *Journal of the American Society for Information Science* 45(9): 645-655.
- Van Rijsbergen, C.J. 1979. *Information Retrieval*. London: Butterworths.

PERTINENT ARTICLES

Chimera, R. & Shneiderman, B. 1994 An exploratory evaluation of three interfaces for browsing large hierarchical tables of contents. *ACM Transactions on Information Systems* 12(4): 383-407.

Newby, G. B. 1992. An investigation of the role of navigation for information retrieval. *ASIS Proceedings '92*: 20-23.

Plaisant, C., Carr, D., and Shneiderman, B. 1995. Image-browser taxonomy and guidelines for designers. *IEEE Software* 12(2): 21-33.