Creating a Navigation Ontology

Raphael Malyankar
Dept. of Computer Science
Box 875406, Arizona State University
Tempe, AZ 85287-5406, USA.
Tel: (602) 965-2804 (v.), (602) 965-2751 (fax)
E-mail: rmm@acm.org
URL: www.eas.asu.edu/~gcss/

Abstract

An effort to prepare an ontology for geospatial information is described. This effort concentrates in the first instance on marine navigational information. This document discusses issues relating to ontology management expected to be encountered during this just-commenced project and our planned approach to dealing with these issues.

Introduction

Even a cursory look at maps and marine charts shows the vast amount of information and knowledge represented in these documents. Further, any map or chart represents information only in context and for a specific purpose (e.g., physical and political maps of the same region), and there are usually many other sources of information that must be concurrently processed by anyone working with a map or chart (such as the navigator of a vessel). The most sophisticated geographical information systems (GIS), ENC (Electronic Navigational Chart) systems, and digital cartographic systems currently available still deal with only very basic geospatial information, for example, routes and waypoints, tide tables, currents, overlays of one kind or another. They are capable of only relatively basic geographic operations such as distance, adjacency, Voronoi diagrams, etc. This paper describes a project on creating an ontology (suitable for automated reasoning) of geospatial knowledge, concentrating in the first instance on marine navigational knowledge, concentrating on some of the ontology management issues involved and our approach to them. The aim of this effort is to allow automated systems to use geospatial information better, so that it is possible for a navigational program to "understand", for example, that shorelines can be crossed by aircraft but not by surface vessels, that tidal tables are important for near shore navigation, that routes may need to stay away from restricted waters, that cutting across shipping lanes should be minimized, etc.

The next section of this paper outlines the research issues arising in this domain. This outline is followed by a description of the sources of ontological knowledge that we expect to use, plans for maintenance and standardization, related research, and problems and limitations encountered so far.

Research Issues

The principles guiding the development of this ontology are as follows: (i) The products of this effort should be of practical use, that is, they should be amenable to automated reasoning or question-answering; (ii) Existing concept definitions, relations, and axioms must be used as far as possible, in order to reduce workload, avoid duplication of effort, enhance acceptability to the user communities, and leverage the vast amount of domain data and knowledge, and ground concepts and inferencing must be grounded in actuality to the maximum extent.

The nature of the domain and the requirement to produce an ontology that can be used for practical reasoning necessitate dealing with the following issues:

Multiple categories of information: Processes, components, mechanisms, and fields, are all present in this domain, necessitating representation with appropriate ontologies. Some information is available only in tabular form, and it will be necessary to find a way to describe these tables, their applicability and contents. It is also necessary to represent information such as geographical features, flow processes such as currents, etc., and also fuzzy objects.

Inaccurate and varying information: Data measurements in this domain are not all accurate to the same extent and will certainly vary with time, both predictably and unpredictably. Predictable variations include such items as tides and unpredictable variations include such items as scour (movement of material on the sea bottom) which will change depth values. Paper charts deal with this by adding warning notes. Allowing for inaccuracy and variations while still retaining practicality is one problem we will need to deal with; we anticipate that the simplest solution, specifying ranges and trends, might not suffice for this domain due to both the nature of the variations and cumulative uncertainty effects for automated inference.

Data Volume: Charts are densely packed with information; even a single chart covering a small region may contain 100's of soundings alone. Tagging this data, and devising representations that are tractable for inferencing with large amounts of data is necessary.

Contexts: Categories of information that are useful (or even available) will be context-dependent. For example, tidal information is unlikely to be needed (or available)

except near shorelines.

Scaling: Details important at one map scale might be irrelevant at another. This is not necessarily the same issue of context. Means of capturing scaling effects and information abstraction must be explored.

Unintuitive, unsophisticated reasoning: In the domain of interest, intuitive reasoning or naive navigation, is often useless because leads to useless results, while 'sophisticated' reasoning is infeasible due to non-availability of data or intractability of necessary computations. What is needed here is either a form of reasoning that lies between 'naivete' and 'sophistication', or a mixture of naive and sophisticated reasoning, or selection based on context. For example, over large distances, the length of the shortest path between two points is calculated using great circle calculations; but this is not necessary over short distances.

The form of available information in this domain often imposes severe restrictions on the possible inferencing. For example, answering a simple question such as "What is the tidal current at $X°$ N. $Y°$ W. at time T"? requires a multi-stage calculation, as follows (assuming the NOAA tide tables):

1. The reference station and subordinate station closest to the point in question are identified. Full tide information is supplied only for a limited number of stations called reference stations. Information about tide 'differences' (differences between times of high and low tide and height relative to a reference station) is available for more than 3000 other locations (subordinate stations). These difference tables also contain information about current speed ratios and average speeds and directions at flood, ebb and slack water.

2. The times of high tide, low tide and slack water closest to the specified time point T and the speeds of the currents at these times are read from the tables for the reference stations.

3. The time differences for the subordinate station are read from the appropriate section of a second table and the times of slack water occurring just before the specified time point T, and the time of the next maximum current are calculated by adding or subtracting the differences from the corresponding times for the reference station. (Difference values in the tables are signed quantities.) The speed ratios and directions of the currents at flood/ebb and slack water are also read from this table.

4. A third table is consulted for an adjustment factor to the speed. The indexes to this table are the interval between slack and maximum current and the interval between slack water and the desired time T. This factor is applied to the speed to give the estimated current velocity at the subordinate station at the required time T. This speed and the bearing looked up in the earlier step may be used as an approximation to the speed and bearing at the desired location.

The above calculation is still an approximation to the desired answer; actual values will depend on local conditions, and indeed more accurate values might be available from other sources such as local marinas. Further, for many purposes, knowledge of current speed and direction at a specific time is not very useful; for some practical applications, it is more useful to know the averages over a period, in which case the above lookups and calculations are repeated for different times and averaged over an interval. An instance of such is the computation of the tidal current vector for a search and rescue plan (USCG 1991).

Sources of Ontological Knowledge

A number of standards efforts for geodata are currently active. The relevant standards and proposed standards contain definitions of terms, which are expected to prove a valuable source of ontological information. One of our tasks is the selection of those sources that are most useful and amenable to computerized formal representations. Some of the sources are outlined below.

Geodata Standards

There is a multitude of standards, proposed standards, and related documents for geodata, such as the Content Standard for Digital Geospatial Metadata(FGDC 1998a), the Spatial Data Transfer Standard (SDTS)(FGDC 1998b), the proposed National Shoreline Data Standard, and other, similar documents from the Federal Geographic Data Committee (FGDC), National Oceanic and Atmospheric Administration (NOAA), etc. Much space in these is devoted to data formatting and field naming issues, necessary for geographical information systems but of limited interest to an ontological engineering project. Ontological information in these mainly consists of term definitions, organized by context (e.g. a shoreline data standard). Taxonomical organization is sometimes but not always explicit from these documents. Based on these documents, we believe taxonomies in this domain are likely to be only a few levels deep.

The SDTS documents (FGDC 1998b) already provide a large number of relevant feature definitions and entity type definitions; it will be possible to leverage this work for ontological engineering purposes. The proposed ontology goes beyond this standard by layering a 'knowledge transfer' capability onto this 'data transfer' standard, by enhancing and formalizing knowledge about relationships between 'features' (geospatial entities), about how these features can be used in automated reasoning systems (e.g., navigational reasoning), by adding microtheories and 'collections' (in the CYC sense) to improve inferencing, and imposing contexts onto this aggregate. The standards have already covered some of the ontological engineering needed, since they contain concept definitions and relationships. For example, Part 2 of the draft Spatial Data Transfer Standard contains the following definitions, selected by starting with HARBOR and following links to some associated concepts such as BREAKWATER,LANE, and (marine) PILOT:

BREAKWATER A structure built to break the force of waves so as to protect a beach, harbor, or other waterfront facility.

HARBOR An area of water where ships, planes or other watercraft can anchor or dock. Also spelled HARBOUR.

LANDING_PLACE A place for loading and unloading passengers or cargo to and from water vessels.

LANE A prescribed course for ships or aircraft, or a strip delineated on a road to accommodate a single line of automobiles; not to be confused with the road itself.

MARINA A harbor facility for recreational craft where supplies, repairs, and various services are available.

PIER A structure built out into the water, usually with its greatest dimension at right angles to the shore, forming a landing place or a place alongside which vessels can lie.

PILOT_WATERS Areas in which the services of a marine pilot are essential.

PORT A landing place provided with terminal and transfer facilities for loading and discharging cargo or passengers, usually located in a harbor.

There are also lists of included terms (defined as non-standard names for entity types or attributes). For example: Anchorage, Artificial harbor, Boat basin, Boat harbor, Harbor of refuge, Haven, Inner harbor, Island harbor, Open berth, Open harbor, and Open roadstead are all included terms for HARBOR; Boom, Groin, Groyne, Jetty, Mole, Sea Gate, Seawall, Wave trap, and Weir trap are included terms for BREAKWATER.

The utility of these concept definitions for creating a taxonomy is immediately obvious. However, these concept definitions were apparently not intended as a complete and coherent ontology, and relationships are not always made explicit; while the standard is a good starting point, we have found it necessary to add a measure of introspection and interpretation to terms, particularly where the relationships between them are not obvious from lexical clues.

Guides and Manuals

Piloting guides and manuals, describing navigation and seamanship techniques are widely available. The "United States Coast Pilot" is a series of publications that provide information to sailors about local navigation hazards, currents, tides, etc. This and similar documentation will also serve to ground the research in practical applications areas.

Information in these documents is often expressed in narrative form and refers to named locations and fuzzy areas. For example, the "Sailing Instructions" for the south coast of Ireland (selected at random from those available) begin as follows:

Plan — This sector describes the S coast of Ireland from Mizen Head to Carnsore point (52° 10′N, 6° 22′W). The descriptive sequence is from W to E.

[Next are general remarks about nature of coastline, prevailing winds, storms, tides and currents, wind effects and other cautions, and traffic rules.]

Coastal features:

1.2 Mizen Head (52°27′N., 9°49′W) is the SW extremity of Cruckaun Island, which is connected to the coast by a narrow neck of land. A ruined tower stands at an elevation of 128m about 0.5 mile NE of Mizen Head and about 0.5 mile NNE of the tower is Mizen Peak, 230m high, the highest hill in vicinity. Mizen Head has been reported to be a good radar target at a distance of 17 miles. A light is shown from a structure on Mizen Head. A radio beacon and a racon are situated on a white building with radio masts, 23m high, close to it, about 0.2 mile NNE of the light structure.

Carrigower Rock lies about 0.5 mile E of Mizen Head and is awash at HW.

Mount Gabriel, 404m in elevation, rises about 12 miles NE of Mizen Head and has conspicuous radar domes near the summit.

There follows information about tide velocity, a caution about tidal race, a remark about the race being particularly dangerous when the wind is opposed to the tidal current, remarks about the velocity decreasing with distance from shore, direction of currents. This is followed by information about a navigational hazard and a cove. Similar information is provided for other features along the coast. A harbor (Cork Harbour) is described in more detail than coastal features, including brief information about cargo facilities, approaches, regulations, and a reference to the World Port Index for more information.

This Sailing Instruction is 15 two-column pages long, plus a single-page map which shows which chart should be used for which part of the shoreline. Charts overlap, since harbor approaches are covered by both small- and large-scale charts. All this for 138 miles of (admittedly well-traveled) shoreline. The magnitude of the task is apparent. Even for this short length of shoreline, it is necessary to reconcile the information about tides and currents with that from tide tables, represent gradations in velocity (which are probably not in the tables), note navigational hazards, aids to navigation (whose descriptions are mostly referent to named locations rather than latitude and longitude), as well as link to harbor and chart information.

Capture of Commonsense Knowledge

Commonsense knowledge for our purpose includes items such as knowing when to use what metric for distance calculations, knowing that a harbor can be used if the draft at the entrance is enough at high water even when an entry at low tide is infeasible, that sailing vessels require a breeze (but not storms) in order to sail, requirements for maneuvering room for vessels, safe distances from rocks and shoals (which will depend on the vessel's draft), and so on. Some of this is intuitively obvious (and this intuition is accurate), some is not obvious to the non-sailor but can be obtained from navigation and seamanship manuals, and some must be obtained through interviews with domain experts. Interviews with personnel from government agencies directly concerned with navigational information and its use are anticipated for the domain expertise related information and also in order to ground our work in practical application areas.

Reuse of Other Ontologies

Geographical information appears to have been dealt with only cursorily in the ontological engineering community to date, though CYC (Lenat 1995) apparently does deal with geographical information to some extent and does contain theories on related information such as terrain, maps, graphic elements for maps, and weather. There has also been some research reported on the representation of fuzzy objects in the knowledge representation and spatial information representation literature (Burrough 1996; Coucelis 1996; Schneider 1996; Cohn & Gotts 1996). However, marine navigation information and knowledge have not attracted attention in ontology work to date.

Maintenance

We anticipate maintenance will be of three kinds:

1. Extensions to the ontology are expected to be performed by knowledge engineers and KR-literate cartographers, who understand formal notation and are comfortable with relatively low-level specification of concepts and relationships.

2. Additions by domain experts, who are unfamiliar with the formal notation used by the ontology, and so will need user-friendly templates for adding knowledge. This implies the creation of methodologies for creation, update, and use, also translation of the ontology between formats that may be used by different planning systems, GIS tools, etc. These methodologies and means for user-friendly enhancement are being investigated.

3. Addition and update of data (e.g., loading of weather and current information), which is not primarily an ontological engineering problem but is still an important part of providing a functional system. This is expected to be an ongoing issue, as the domain is extremely dynamic and information changes at varying rates (over time scales of hours to years).

Standardization

For our particular problem, compatibility with existing geodata standards is essential. A transition of the ontology into a future standard is also a possibility, ensuring compatibility with existing and proposed data standards, and taking full advantage of standards work in the context of the National Spatial Data Infrastructure (NSDI) and elsewhere. As for standardization of representation, we expect to use existing ontological formalisms for representation and inferencing. The questions of tractability and completeness in this domain have yet to be explored.

Scoping and Modularity

Information in this domain divides naturally along contextual and location-specific grounds, expected to lead to the following treatments:

1. Defining the scope and application of research into geospatial ontologies, including the definition of mi-

crotheories within such ontologies (such as a microtheory for inshore navigation, one for air navigation, etc.).

2. Representations of data such as tide tables, water temperatures, currents, the location-specificity and time-specificity of such data, and its use in inferencing, information retrieval, and planning.

We expect to have to define different microtheories in different geospatial contexts, for example, land navigation is different from maritime navigation, and air navigation is different from both; further, maritime navigation is different on the high seas than in near-shore waters. These differences are qualitative in nature: for example, tides need not be considered in high-seas contexts, though weather and distance are very important; but close to land, tides are extremely important, and so are obstacles and wind and current speed and directions. However there is still significant overlap between land, sea, and air navigation, for example, the need for great circle routes over long distances (common to air and high-seas navigation), path planning with waypoints, and political boundaries.

Integration

Integration with AI applied research into planning and information retrieval is expected. An important consideration for the problem is integration with current and future geographical information systems, in the senses of information representation, layering of contexts over one another as well as over low-level data, and integration of an ontological standard with data and metadata standards.

Application

Our initial grounding application planned is a Web-based question-answering system that provides plots of routes between two points and also information about navigational hazards that might be encountered on that route and other information that the master of a vessel sailing it should be aware of. Route plotting and tidal information is already available in commercial ENC (electronic navigation chart) systems, but retrieval appears to be keyed mainly to geolocation. Sailing directions for some parts of the world have already been made available online, and there are plans to make retrieval of these more intelligent, apparently again by keying to geolocation. Our aim is to make answers relevant and context-sensitive, so that, for example, a pleasure boat will not be provided information about cargo unloading facilities at the destination, but will get information about marinas, while the master of a cargo ship will get the cargo unloading information, and information about draft and pilotage rules. Evaluation of this system and ultimately of the ontology will be made based on the relevance and completeness of answers to queries within the system's expertise and the recognition (and 'graceful handling') of those that are not entirely within its domain.

Related Research

The most relevant research initiatives and conferences appear to be mainly in geography and cartography, namely

51

Project Varenius and related efforts of the National Center for Geographical Information and Analysis (NCGIA)(Mark, Egenhofer, & Hornsby 1997; Peuquet, Smith, & Brogaard 1998), the series of Conferences on Spatial Information Theory, etc.

Related work by other groups falls into three major fields: ontological engineering and knowledge representation, AI planning, and cartography, GIS, and geospatial information representation.

Ontological Engineering and Knowledge Representation

The CYC project (Guha & Lenat 1990; Lenat 1995) (of MCC and later Cycorp) contains some geospatial information in its "Geography" collection and there is apparently a certain amount of current interest in using Cyc in GIS. An examination of the information publicly available from Cycorp indicates that the geospatial relations may be limited to high-level information retrieval rather than navigation. Neither the list of applications for CYC mentioned by Cycorp, nor the projects currently listed as being part of the DARPA High-Performance Knowledge Base (HPKB) project appear to include detailed geospatial knowledge, still less the kind of geospatial and navigational knowledge envisaged in this proposal, even though rudimentary geospatial concepts such as "body of water" are used in those efforts. Further, relating whatever geospatial knowledge is included in CYC and the HPKB projects to existing and future standards for charts, and geospatial knowledge standards does not appear to be envisaged, as the intent of these projects (CYC, HPKB, etc.) is not primarily geospatial knowledge representation.

One major difference from the Cyc project is that it deals with common-sense reasoning — where results could be wrong, though consistent — while our project deals with representing geospatial knowledge, that is, scientific and cartographic, well-defined. 'correct', internally consistent knowledge for relatively rigorous reasoning and use (as compared to common-sense reasoning).

Other knowledge representation work dealing with crisp and fuzzy objects and boundaries between objects (Borgo, Guarino, & Masolo 1996b; 1996a; Burrough 1996; Cohn & Gotts 1996; Frank 1996; Smith & Varzi 1997) spatial reasoning, representation of fields and flows, and the representation of time and temporal dependencies will also be relevant to the proposed work.

Artificial Intelligence Planning Research

Route planning and terrestrial navigation research has been mostly land-based to date, though some work on navigation for autonomous submersibles has been done. There is a significant body of robotics research in land navigation; notable projects in this field include the CMU work on road-following, a significant body of work on maps for autonomous robots, etc., NASA work on autonomous planetary rovers, notably the Mars Sojourner vehicle, and others too numerous to describe here. However, much of this work has been directed at relatively small-scale maps of local areas, and building cognitive representations of the environment. These representations are necessarily somewhat sparse, due to the current state of computational resources and techniques and the early phase of such work. We expect techniques and lessons learned will transfer to our ontology project.

Cartography, GIS, and Geospatial Knowledge Representation

There is currently an active effort by the Federal Geographic Data Committee (FGDC) to construct standards for geographic data and meta-data. Some standards have already been published (FGDC 1998a; 1998b). These standards describe data formats and data elements for digitized maps and charts, features, graphic elements and standards, annotations describing data reliability, source, and resolution, and other kinds of *data*, as opposed to *knowledge*. Current GIS's generally work at a relatively low level in comparison to AI planning systems — in many cases, they do not even work at the symbol level, let alone reaching the knowledge level, in the sense in which these levels are described by Newell (Newell 1982).

Research into geospatial information representation is very relevant to the proposed project, but too voluminous to describe here in any detail. Some non-exhaustive examples of the kind of work we believe to be relevant here are are the NCGIA specialist meetings (Mark, Egenhofer, & Hornsby 1997; Peuquet, Smith, & Brogaard 1998), representation of object extent and boundaries(Coucelis 1996; Frank 1996; Smith & Varzi 1997), representation of fields and flows (Kavouras 1996), and pedagogic texts dealing with geographic knowledge and its common-sense interpretations and use.

Current Status

Our current efforts are mainly concerned with the analysis of existing and proposed standards for digital geospatial data representation, particularly the Spatial Data Transfer Standard mentioned earlier, and 'ontologization' of these standards through lexical analysis and human intervention for addition of relationships, axioms, definition of collections and microtheories. This is combined with exploratory development of the question-answering Web-based system described earlier. We are also engaged in defining reasoning methods required for grounding purposes and that can be used in the question-answering system.

Limitations and Problems Encountered

One significant problem we have encountered so far in our exploratory effort (that is not apparent in existing efforts on developing geographic ontologies) is the volume of data that must be made available in the system, which is needed in order to produce certain efforts. For example, tide tables give tide information for every day of the year, and this information is location-specific and also changes from year to year; this alone requires significant storage of data even for

relatively shallow inferencing procedures. The second significant problem is transforming the large volume of narrative information (as in the example from Sailing Instructions provided earlier in this paper) into a suitable formalism.

Summary

This paper has described issues arising in the creation and formalization of a marine navigation ontology, research issues in this domain, the sources of ontological knowledge and means of extracting this ontological knowledge, and problems encountered to date. One question arises, based on our experiences and discussions of issues above: Is this a database and information retrieval problem, or a ontological engineering and knowledge representation problem? Prior treatments and commercial systems treat it as a database and information retrieval issue. We believe, however, that it is primarily an AI and knowledge representation problem, because though the volume of information is large and information retrieval plays an important role, the critical factor in using this information is human information-processing of the retrieved information and data, for example knowing when and why to take currents and tide effects into account, what parts of sailing instructions are important at a point of time and what parts will be relevant in the near future, the question of deciding in what context reasoning should be done. Our demonstration question-answering system is, in a small way, an attempt to reproduce some of the results of the thought processes of a human engaged in the task of navigation.

Acknowledgements

References

Borgo, S.; Guarino, N.; and Masolo, C. 1996a. A pointless theory of space based on strong connection and congruence. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR '96)*, 220–229.

Borgo, S.; Guarino, N.; and Masolo, C. 1996b. Stratified ontologies: The case of physical objects. In *ECAI-96 Workshop on Ontological Engineering*.

Burrough, P. A. 1996. Natural objects with indeterminate boundaries. In *Geographic Objects with Indeterminate Boundaries (GISDATA2)*. London: Taylor and Francis Ltd. chapter 1, 3–28.

Cohn, A. G., and Gotts, N. M. 1996. Representing spatial vagueness: A mereological approach. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR '96)*, 230–241.

Coucelis, H. 1996. Towards an operational typology of geographic entities with ill-defined boundaries. In *Geographic Objects with Indeterminate Boundaries*. London: Taylor and Francis Ltd. chapter 3, 45–56.

FGDC. 1998a. Content standard for digital geospatial metadata. Federal Geographic Data Committee, U. S. Geological Survey. FGDC-STD-001-1998.

FGDC. 1998b. Spatial data transfer standard. Federal Geographic Data Committee, U. S. Geological Survey. Proposed standard.

Frank, A. U. 1996. The prevalence of objects with sharp boundaries in GIS. In *Geographic Objects with Indeterminate Boundaries*. London: Taylor and Francis Ltd. chapter 2, 29–40.

Guha, R. V., and Lenat, D. B. 1990. *Building large knowledge-based systems : representation and inference in the Cyc project*. Reading, MA: Addison Wesley.

Kavouras, M. 1996. Geoscience modelling: From continuous fields to entities. In *Geographic Objects with Indeterminate Boundaries*. London: Taylor and Francis Ltd. chapter 21, 313–324.

Lenat, D. B. 1995. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM* 38(11):32–38.

Mark, D. M.; Egenhofer, M. J.; and Hornsby, K. 1997. Formal models of commonsense geographic worlds: Report on the specialist meeting of research initiative 21. Technical Report 97-2, National Center for Geographic Information and Analysis, San Marcos, TX.

Newell, A. 1982. The knowledge level. *Artificial Intelligence* 18(1):87–127.

Peuquet, D.; Smith, B.; and Brogaard, B. 1998. The ontology of fields: Report of a specialist meeting held under the auspices of the varenius project: Panel on computational implementations of geographic concepts. Technical report, National Center for Geographic Information and Analysis, Bar Harbor, ME.

Schneider, M. 1996. Modelling spatial objects with undetermined boundaries using the Realm/ROSE approach. In *Geographic Objects with Indeterminate Boundaries*. London: Taylor and Francis Ltd. chapter 10, 141–152.

Smith, B., and Varzi, A. C. 1997. Fiat and bona fide boundaries. In *Spatial Information Theory, A Theoretical Basis for GIS: COSIT '97, Laurel Highlands, Pennsylvania*, 103–119. Springer-Verlag.

USCG. 1991. *National Search and Rescue Manual (Planning Handbook)*. United States Coast Guard, Washington, DC.