Ontology of Common Sense Geographic Phenomena: Foundations for Interoperable Multilingual Geospatial Databases

David M. Mark, Barry Smith, Berit Brogaard-Pedersen
National Center for Geographic Information and Analysis
University at Buffalo
Buffalo, New York, USA
e-mail: dmark@geog.buffalo.edu

Abstract

Information may be defined as the conceptual or communicable part of the content of mental acts. The content of mental acts includes sensory data as well as concepts, particular as well as general information. An information system is an external (non-mental) system designed to store such content. Information systems afford indirect transmission of content between people, some of whom may put information into the system and others who are among those who use the system. In order for communication to happen, the conceptual systems of the originators and users of the information must be sufficiently similar. A formal conceptual framework that can provide the basis for exchange of information is termed an ontology. In its most fundamental form, ontology studies the most basic constituents of reality. Traditionally, ontology seeks to reflects structures that are independent of thought and cognition. The term ontology is used more broadly in artificial intelligence and software engineering, to refer to the conceptual basis for an information system.

As part of our investigations of the ontological foundations of geographic information science (Smith and Mark, 1998, 1999), we are studying the concepts that every-day people use when thinking and communicating about the geographic world, the categories used in people's naive geographies (Egenhofer and Mark, 1995). This paper is based on an experimental protocol that elicits examples of categories from subjects. This protocol follows one used by Battig and Montague (1968) to elicit such category examples, which they called norms; we piloted this protocol for geographic categories last year, and reported some preliminary results (Mark at al., 1999). In this protocol, subjects are given 30 seconds to list examples for a phrase such as "a kind of geographic feature". They were then told to stop, turn the page, and list examples of some other category, again for 30 seconds, until eleven categories had been exemplified. In this paper, we will report results for just two sets of these categories. Each subject was asked to give examples for one of the following five phrases:

1) a kind of geographic feature
2) a kind of geographic object
3) something geographic
4) a geographic concept
5) or something that could be portrayed on a map

Each also was asked for the negation of the same phrase, such as "a kind of non-geographic feature" or "something that could not be portrayed on a map."

A total of 263 undergraduate students at the University at Buffalo participated in the experiment that resulted in the data for the English language. The subjects listed 308 different words or phrases, but 35 terms made up 72 percent of the examples. Most analysis will focus on these 35 terms.

Keywords: geographic information science, geographic database, semantic interoperability, ontology, multilingual, categories

Results

Examples per Subject

The mean number of examples give per subject is probably an indication of the naturalness or familiarity of the category. This measure varied substantially across the different phrasings, with the highest average being 8.21 examples per subject for 'mappable'. Also highwas 'a kind of geographic feature', which produced an average of 7.15 examples per subject for 'feature'. In contrast, 'a kind of geographic concept' produced an average of only 5.15 examples per subject for 'concept', less than 2/3 as many as for 'mappable'. 'A kind of geographic object' also was low, at 5.48 examples per subject for 'object'. 'Something geographic', perhaps the most generic way to pose the question, produced the most intermediate result, 6.17 examples per subject. It seems likely that 'something that could be portrayed on a map' is a more familiar idea than 'geographic' or 'geography', not surprising considering the near-ubiquity of maps U.S. society, and the low profile of the discipline of geography in this country. It will be interesting to see if this trend is true in other countries.

In order to better define these categories, we also asked subjects to give examples of 'opposite' categories: 'something that is not geographic', 'a kind of object that is not geographic', 'a non-geographic concept', 'something that could not be shown on a map', and 'a kind of feature that is not geographic'. In this case, 'object' had the highest frequency per subject, at 4.86, with 'something non-greographic' a close second, at 4.71. The two forms that produced the most examples as cited above produced the fewest examples when stated in the negative. Subjects were able to list only an average of 2.54 examples each for non-geographic features, and 3.04 examples each for 'something that could not be shown on a map'

Content of the Categories

Four terms, namely mountain, river, lake, and ocean, were commonly listed in response to all versions of the question. However, there were important differences. First, almost all responses to "a kind of geographic feature"denoted physical features of the earth, confirming a preliminary result reported by Mark at al. (1999). On the other hand, "something that could be portrayed on a map" elicited many artificial features, such as city, road, and country. Also of note, "a kind of geographic object" elicited several kinds of non-geographic things that are used to represent or measure geographic things, in particular map, globe, compass, and atlas. The most generic form of the question, 'something geographic', produced intermediate results. Mobility and size appear to be key factors in deciding what qualifies as geographic, with the exception of manipulable artifacts used for geographic purposes. Among natural objects of geographic scales, the most frequent responses were a mix of shape-based landforms such as mountain (151), water courses such as river (129), and water bodies such as lake (102) and ocean (95).

For many of the phrasings for the negative (non-geographic) categories, predominant examples were relatively small (though not tiny) and moveable. For example, 'non-mappable' was lead by people (13 cases), cars (8), animals (7) being more frequent that houses (6). results for non-geographic were similar, but the top two examples, cars (16) and people (8), were reversed. Two other categories, 'object' and 'feature', produced somewhat different results. 'Non-geographic object' was led by car (12), but next were furniture items such as chair (8), desk (7), and table (6), as well as some much smaller objects such as pen (8) and pencil (7). On the other hand, 'non-geographic feature' produced non-moveable objects such as building (13) and house (8) as most frequent answers, with people mentioned by 5 subjects and cars by 4. Examples for 'non-geographic concept' showed little consensus, with led by money (4), time (4), space (4), and love (2).

Future Work

Plans are underway to replicate the experiment in Finnish, French, Spanish, German, Danish, Japanese, and Mandarin. Results of the experiment for Finnish and perhaps other European languages will be available by May 2000. They will be analyzed, compared to the results for English, and presented at the AGILE 2000 meeting if this proposed paper is accepted for presentation at the meeting.

References

Battig, W. F., and Montague, W. E., 1968 "Category Norms for Verbal Items in 56 Categories: A Replication and Extension of the Connecticut Norms," Journal of Experimental Psychology Monograph, 80, No. 3, Part 2, pp. 1-46.

Egenhofer, M. J., and Mark, D. M., 1995. Naive Geography. In Frank, A. U. and Kuhn, W., editors, Spatial Information Theory: A Theoretical Basis for GIS, Berlin: Springer-Verlag, Lecture Notes in Computer Sciences No. 988, pp. 1-15.

Mark, D. M., Smith, B., and Tversky, B., 1999. Ontology and Geographic Objects: An Empirical Study of Cognitive Categorization. In Freksa, C., and Mark, D. M., editors, Spatial Information Theory: A Theoretical Basis for GIS, Berlin: Springer-Verlag, Lecture Notes in Computer Sciences, pp. 283-298.

Smith, B., and Mark, D. M., 1998. Ontology and Geographic Kinds. in T. K. Poiker and N. Chrisman (eds.), Proceedings. 8th International Symposium on Spatial Data Handling (SDH98), Vancouver: International Geographical Union, 1998, 308-320.

Smith, B., and Mark, D. M., 1999. Ontology with Human Subjects Testing. American Journal of Economics and Sociology, v. 58(2): 245-272.