A GEOGRAPHICAL DATABASE SYSTEM

Willem van Biljon
National Research Institute for Mathematical Sciences
CSIR, P O Box 395, PRETORIA, 0001, South Africa

ABSTRACT

A geographical database that forms the kernel of a geographical information system (GIS) is described. The database facilitates topologically structured spatial attributes as well as a poset (partially ordered set) classification and non-spatial attributes scheme. The database provides the capability of attaching more than one set of spatial attributes to a single feature to enable the GIS to perform semi-automatic generalization.

INTRODUCTION

The National Research Institute for Mathematical Sciences (NRIMS) of the Council for Scientific and Industrial Research (CSIR) undertook research into geographical information systems (GIS) by developing a prototype computer-assisted cartography system. The development and use of this system has led to a better understanding of the issues involved, and NRIMS is currently designing and implementing a complete GIS.

The major requirements for the GIS are that, in addition to the cartographic capabilities, it should provide fully interactive query facilities, both graphically and alphanumerically. This paper describes the design of the geographical database that forms the kernel of the GIS and in which all feature attributes are stored. Both the objectives for and the resulting design of the database are discussed.

OBJECTIVES

A number of objectives were set concerning the management of attributes in the database. These are the result of two factors. Firstly, a number of requirements were identified during the use of a prototype cartographic system, and secondly, some practical objectives were set in accordance with theoretical objectives of the system. Of importance in this regard are those requirements that relate to completeness of the database. That is, its capability to answer any metric, topological or geographical query. These requirements have been examined by (White, 1984). A summary of the objectives follows.

Data integrity, consistency and reduced redundance

These requirements are usually found in any database design.

Maintenance of data topology

To enable the database to contain enough information to answer all topological and geographical queries it is essential that the

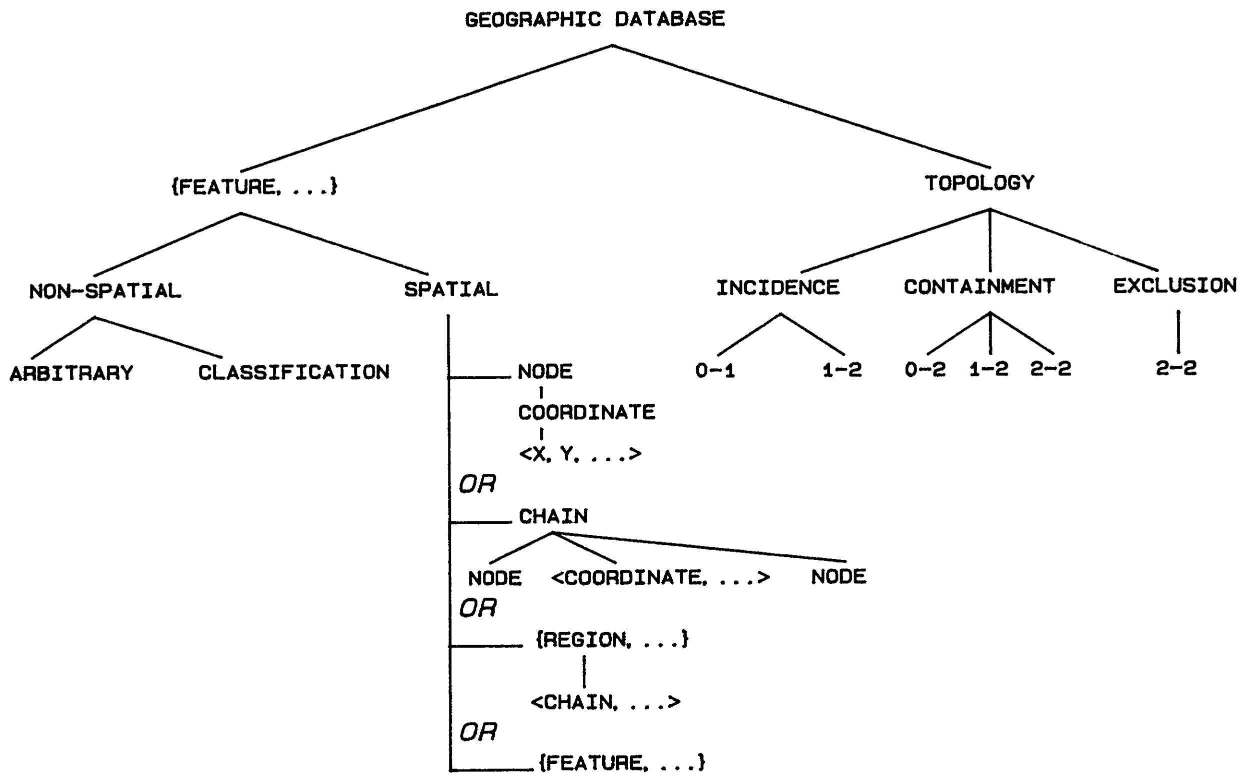

Figure 1: Structure of the Database

topology of the data should be maintained. This information should be independent of spatial attributes since the topology is a function of structure rather than spatial position.

Efficiency of spatial manipulation

Large volumes of spatial information are stored in a GIS. It is essential that efficient data structures should be provided to allow fast interactive manipulation and display of information.

Multiple spatial data sets per feature

Provision must be made to associate a number of different spatial data sets for a single feature. This will allow the digitization of different views of a feature, and may be used to associate data sets that have varying amounts of detail with a feature.

Decomposition of features into sets

It should be possible to construct features from sets of other features. This has two benefits. Firstly, it allows sets of features to be associated, and hence to have a semantic connection. Secondly, this facility, together with multiple data sets, allows the system to perform generalization.

Management of non-spatial attributes in a feature-dependent manner

Features of different types (or classes) have different attributes. A mechanism must be provided to manage these attributes in a feature-dependent manner.

The database was designed with these objectives as primary concerns.

CONSTRUCTION OF THE DATABASE

Figure 1 gives an informal graphical overview of the design of the database. The tree structure depicts each entity in the database, as well as the components of which it consists. Curly brackets denote sets of entities, whereas angle brackets denote arrays of entities, that is, order is important.

The database consists of two components: a set of features and a collection of topologies on the spatial attributes of the features.

Features

A feature represents a natural, man-made or abstract object that exists on the surface of the earth. A feature is described by two types of attributes: spatial and non-spatial. Each feature has an array (or vector) of spatial attributes associated with it, as well as a single set of non-spatial attributes.

Spatial attributes. Each spatial attribute of a given feature is either a point attribute (and we speak of a **point feature**), a line attribute (**line feature**) or an area attribute (**area feature**). Finally, the spatial attribute may be a set of other features (and we speak of a **compound feature**).

A point feature maps directly down to a node, which in turn consists of a coordinate in n-space, usually either on the projection plane or on the spheroid.

A line feature maps down to a list of chains. A chain is a list of coordinates terminated at both ends by nodes.
An area feature maps to a set of regions. Each region is enclosed by a closed boundary constructed from a list of chains.

The nodes terminating a chain may be shared by different chains and point features. Likewise, chains may be shared by different line features and region boundaries. Finally, area features may share different regions. This data sharing addresses the objective of reduced data redundance, and simplifies consistency checking in the database.

Multiple spatial attributes for a single feature are based on the philosophy that spatial attributes may differ according to use; these attributes do not necessarily attempt to reflect a single model of reality. The structure within each element of the vector of spatial attributes is similar to structures described in the literature (Guptill, 1986 and Peuquet, 1984).

This structure enables one to create a feature that is, for example, a point feature (when examining spatial attribute 1, say), a line feature (spatial attribute 2), an area feature (spatial attribute 3), or a set of features (spatial attribute 4). An example might be an airport digitized as a point feature at a small scale, as a line feature showing the main runway at a larger scale, or as an area feature at an even larger scale. Finally, the feature may be defined by its constituent runways, hangars, control towers, etc., at a very large scale.

Although this use of a vector of spatial attributes is not the only use (another example may be different views of a feature as digitized from photographs taken in different wavelengths), it is the motivating use for having a vector of spatial attributes. This vector is thus often referred to as detail levels.

Associating each detail level with a scale range provides a mechanism for displaying only an appropriate amount of detail at a certain scale, that is, generalization.

Efficient spatial attribute manipulation is obtained by imposing a quad-tree index on the spatial attributes. The set of quad-tree leaves containing each node, chain and region is computed according to a scheme described by Abel and Smith (1982). Using only the leaves of the quad-tree imposes some storage overheads, but decreases the response time to spatial queries.

<u>Non-spatial attributes</u>. Four requirements must be satisfied by the non-spatial attributes of a feature. These attributes must provide:

1. a classification scheme for feature coding;
2. the type of a feature at each detail level;
3. the allowable set members of a compound feature;
4. definitions of other descriptive attributes.

These requirements may be met by constructing a poset of classes and subclasses, each feature classification being a complete path name in the poset. Figure 2 gives an example of a poset defining some classes for cultural features. The vertices are distinct feature classes, and the arcs denote a major class/subclass relation. A path such as "cultural features/aviation/airport" may be codified as an integer and used as a feature classification. Some vertices in the network may not be complete classifications (for example, "cultural features/aviation"), and hence a distinction is made between these two types of vertices. If a vertex completes a classification then it has associated with it the definition of the feature types at each detail level of a feature of this class, as well as an indication of which compound features the feature may be a part.

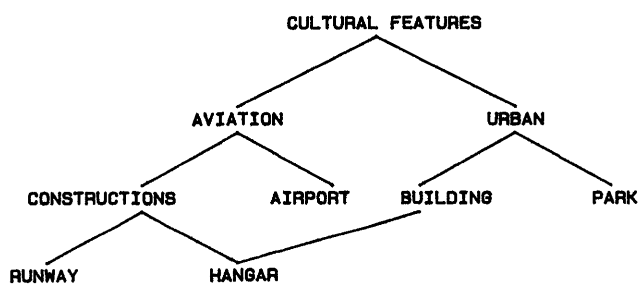

Figure 2: Example subset of the classification of 'Cultural Features'

Finally, each vertex in the poset may have a number of other associated attributes. Subclasses (that is, classes lower down in the poset) inherit all attributes of their major classes. All classes have the same virtual class as a major class. This class associates all attributes common to all features, for example feature name. The attributes may have **computed** or **constraint** clauses defined on them, stating the interactions of the attributes. An example of how some of the classes in Figure 2 may be coded follows. Comments are written following a double hyphen.

```
Cultural Features refines Class   -- No further information of this
end.                              -- subclass is required.  It has no
                                  -- other spatial or non-spatial
                                  -- attributes.

Aviation refines Cultural Features
end
```

```
Airport refines Aviation;           -- This completes a feature
  Spatial: point at 1;              -- classification and thus includes
           line at 2;               -- information on the spatial
           area at 3;               -- attributes for each detail level.
           set at 4;                -- The computed clause specifies how
  Non-spatial: area : numeric;      -- area may be calculated from the
    computed at 4 :                   -- area attributes of elements of
                                    the
      sum (airport.4.set = area);-- composite feature.
end.

Hangar refines Constructions, Buildings; part of Airport;
  Spatial: point at 1,2,3;           -- The part of clause specifies
           area at 4;                -- that a hangar may be included
  Non-spatial: area : numeric;       -- in the set of airport
end.                                 -- composite spatial attributes.
                                     -- The area attribute is
                                     -- referred to by the computed
                                     -- clause in the airport class.
```

Implementation of this scheme borrows concepts from object-oriented languages.

TOPOLOGY

The topology of a GIS contains information about the structure of spatial attributes and is independent of the form of the attribute. Three types of topological relationships are discussed: incidence, containment and exclusion.

Incidence

Incidence information between nodes and chains (0-1 incidence) and between chains and regions (1-2 incidence) is kept. Node-region (0-2) and chain-chain (1-1) incidence may be derived from these owing to the transitive nature of the incidence relation. Some of this information is already available from the construction of the database (specifically chain→node and region→chain relations). The other information (node→chain and chain→region) is coded explicitly. The use of incidence relations has been discussed extensively by White (1984).

Containment

Containment information is required to answer geographical queries. The relations required are nodes inside regions (0-2 containment), chains inside regions (1-2 containment) and regions inside regions (2-2 containment). Although one of the design objectives is to keep all topological information separate from the spatial attributes, this can be relaxed for containment if one assumes that no projection (or map deformation) will map a node or chain either into or out of a region. This assumption is not unrealistic since most projections are continuous functions under the normal distance metric in n-space. Under this assumption containment can always be calculated, at the obvious expence of response time.

Exclusion

For metric queries concerning area features, and in order to render regions correctly during cartographic representation, it is necessary to keep information about regions excluded from other regions. Since nodes and chains have no area, they need not be considered for the exclusion relation. The maintenance of this relation results in an extension to the definition of area features. Each area feature, in addition to a number of included regions, also has a number of excluded regions for each included region.

In contrast to containment, exclusion cannot be calculated since it is a semantic concept, not a spatial one.

Clearly, the topology on a set of spatial attributes is dependent on the actual types of the spatial attributes of features (that is, point, line or area). This means that it is only sensible to make topological queries on a single detail level, since no topology is defined across detail levels. This constraint may be enforced by the user interface of the GIS.

CONCLUSION

A number of objectives for a geographic database and a design that meets these objectives have been described. A prototype of the database using a simplified strictly tree structured hierarchical classification scheme has been implemented using a commercial database management system.

Further work will include the implementation of the complete poset scheme for the management of non-spatial attributes, and the extension of the concepts to include multi-user access without compromising the integrity of features or topology.

REFERENCES

Abel, D J and Smith, J L, 1982. "A Data Structure and Algorithm Based on a linear Key for a Rectangle Retrieval Problem", Computer Vision, Graphics, and Image Processing 24, 1 - 13.

Guptil, S C, 1986. "A New Design for the US Geological Survey's National Digital Cartographic Data Base", Procedings of Auto Carto London, Vol 2, 10 - 18.

Peuquet, D J, 1984. "A Conceptual Framework and Comparison of Spatial Data Models", Cartographica, Vol 21 No 4, 66 - 113.

White, M S, 1984. "Technical Requirements and Standards for a Multipurpose Geographic Data System." The Americal Cartographer, Vol 11, No 1, 15 - 26.