Chapter II
Querying GML:
A Pressing Need

Jose E. Córcoles
Castilla La-Mancha University, Spain

Pascual González
Castilla La-Mancha University, Spain

ABSTRACT

As a database format, XML (GML by extension) can be queried. In order to do this, we need a query language (of general use) to retrieve information from an XML document. Nevertheless, it is necessary to enrich the query language over XML features with spatial operators if we wish to apply it over spatial data encoded with GML. Otherwise, these query languages could only be used to query alphanumeric features of an XML document and not, for example, the topological relationship between two spatial regions. Today, there is a large set of query languages over XML. These query languages are different with respect to syntax, available operators and environment of applicability. However, they share the same features, that is, features of query languages over semi-structured data. With respect to GML, from the literature, it is known that four GML query languages have been proposed. The following chapter briefly describes these query languages over GML.

INTRODUCTION

Today, the World Wide Web speaks a common language named eXtendible Markup Language (**XML**) (W3C, 2005). **XML** is a W3C-recommended general-purpose markup language for creating special-purpose languages, capable of describing many different kinds of data. Although XML is a recent technology, its structure and its original aim is the recovery of an ancient proposal named Standard Generalized Markup Language (SGML), which dates back to the 1970s.

XML is mainly used on the Web as an exchange format, that is, its primary purpose is to facilitate the sharing of data across different systems. This use solves the syntactic heterogenic, as both sources use the same structure in order to represent their data. However, XML has been extended in many more applications, for example, configuration files, databases, and so forth.

XML has several features that make it well-suited for its purpose. Some of these are: it is in both human and machine-readable format; it supports Unicode, allowing almost any information in any human language to be communicated; the ability to represent the most general computer science data structures: records, lists and trees; the self-documenting format that describes structure and field names as well as specific values; the strict syntax and parsing requirements that allow the necessary parsing algorithms to remain simple, efficient, and consistent.

Although XML has a general purpose, there is a wide range of specific-purpose languages based on XML. For example, Geography Markup Language (**GML**), Resource Description Framework (RDF), Scalable Vector Graphics (SVG), Mathematical Markup Langauge (MathML) and Virtual Reality Markup Language (VRML). All of them are defined in a formal way, allowing programs to modify and validate documents in these languages without prior knowledge of their form.

In particular, GML, defined by Open Geospatial Consortium (Open Geospatial Consortium, 2003), plays an important role in spatial systems in general and in geospatial systems in particular. GML can be defined as follows: "An XML encoding for the transport and storage of geographic information, including both the spatial and non-spatial properties of geographic features".

However, a formal definition given by the GML specification is: "Geography Markup Language is an XML grammar written in XML Schema for the modelling, transport, and storage of geographic information".

The key concepts used by GML to model the world are drawn from the OpenGIS Abstract Specification and the ISO 19100 series.

GML provides a variety of objects for describing geography including features, coordinate reference systems, geometry, topology, time, units of measure, and generalized values.

A geographic feature is "an abstraction of a real world phenomenon; it is a geographic feature if it is associated with a location relative to the Earth". So a digital representation of the real world can be thought of as a set of features. The state of a feature is defined by a set of properties, where each property can be thought of as a {name, type, value} triple.

The number of properties a feature may have, together with their names and types, are determined by its type definition. Geographic features with geometry are those with properties that may be geometry-valued. A feature collection is a collection of features that can itself be regarded as a feature; as a consequence of this a feature collection has a feature type and thus may have its own distinct properties, in addition to the features it contains.

XML AS A DATABASE: QUERYING XML

Undoubtedly, an interesting advantage of XML is that it can be used as a database, even though an XML document is a database only in the strictest sense of the term, i.e., it is a collection of data. In many ways, this makes it no different from any other file.

As a database format, XML has several advantages. For example, it is self-describing (the markup describes the structure and type names of the data, although not the semantics), it is portable (Unicode), and it can describe data in tree or graph structures.

These properties make it possible to open a new set of XML applications, all of them involving storage and retrieval of information represented by XML.

One example of XML as a database is a restaurant catalog. It could be defined with alphanumeric features, e.g., name, phone number, address,