Chapter XLIII
GeoCache:
A Cache for GML Geographical Data

Lionel Savary
PRiSM Laboratory, France

Georges Gardarin
PRiSM Laboratory, France

Karine Zeitouni
PRiSM Laboratory, France

ABSTRACT

GML is a promising model for integrating geodata within data warehouses. The resulting databases are generally large and require spatial operators to be handled. Depending on the size of the target geographical data and the number and complexity of operators in a query, the processing time may quickly become prohibitive. To optimize spatial queries over GML encoded data, this chapter introduces a novel cache-based architecture. A new cache replacement policy is then proposed. It takes into account the containment properties of geographical data and predicates, and allows evicting the most irrelevant values from the cache. Experiences with the GeoCache prototype show the effectiveness of the proposed architecture with the associated replacement policy, compared to existing works.

INTRODUCTION

The increasing accumulation of geographical data and the heterogeneity of Geographical Information Systems (GISs) make difficult efficient query processing in distributed GIS. Novel archi-tectures (Boucelma, Messid, & Lacroix, 2002; Chen, Wang, & Rundensteiner, 2004; Corocoles & Gonzalez, 2003; Gupta, Marciano, Zaslavsky, & Baru, 1999; Leclercq, Djamal, & Yétongnon, 1999;Sindoni, Tininini, Ambrosetti, Bedeschi, De Francisci, Gargano, Molinaro, Paolucci, Patteri,

& Ticca, 2001; Stoimenov, Djordjevic-Kajan, & Stojanovic, 2000; Voisard & Juergens, 1999; Zhang, Javed, Shaheen, & Gruenwald, 2001) are based on XML, which becomes a standard for exchanging data between heterogeneous sources. Proposed by OpenGIS (2003), GML is an XML encoding for the modeling, transport, and storage of geographical information including both the spatial and non-spatial fragments of geographical data (called features). As stressed in (Savary & Zeitouni, 2003), we believe that GML is a promising model for geographical data mediating and warehousing purpose.

By their nature, geographical data are large. Thus GML documents are often of important size. The processing time of geographical queries over such documents in a data warehouse can become too large for several reasons:

1. The query evaluator needs to parse entire documents to find and extract query relevant data.
2. Spatial operators are not cost effective, especially if the query contains complex selections and joins on large GML documents.

Moreover, computational costs of spatial operators are generally more expensive than those of standard relational operators. Thus, geographical queries on GML documents raise the problem of memory and CPU consumption. To solve this problem, we propose to exploit the specificities of a semantic cache (Dar, Franklin, Jonsson, Srivastava, & Tan, 1996) with an optimized data structure. The proposed structure aims at considerably reducing memory space by avoiding storing redundant values. Furthermore, a new cache replacement policy is proposed. It keeps in cache the most relevant data for better efficiency.

Related works generally focus on spatial data stored in object-relational databases (Beckmann, Kriegel, Schneider, & Seeger, 1990). The pro-posed cache organizations are better suitable for tuple-oriented data structures (Brinkhoff, 2002). Most cache replacement policies are based on Least Recently Used (LRU) and its variants. Other cache replacement policies proposed in the literature (Arlitt, Friedrich, Cherkasova, Dilley, & Jin, 1999; Cao & Irani, 1997; Lorenzetti & Rizzo, 1996) deal with relational or XML databases, but have not yet investigated the area of XML spatial databases.

The rest of the chapter is organized as follows: The second section gives an overview of related works. In the third section we present our cache architecture adapted for GML geographical data. The fourth section discusses the inference rules of spatial operators and presents an efficient replacement policy for geographical data considering inference between spatial operators. The fifth section shows some results of the proposed cache implementation and replacement policy. Finally, the conclusion summarizes our contributions and points out the main advantages of the proposed GML cache-based architecture.

RELATED WORKS

Cache Replacement Policy

In the literature, several approaches have been proposed for cache replacement policy. The most well known is the Least Recently Used (LRU) (Tanenbaum, 1992). This algorithm replaces the document requested the least recently. Rather at the opposite, the Least Frequently Used (LFU) algorithm evicts the document accessed the least frequently. A lot of extensions or variations have been proposed in the context of WWW proxy caching algorithms. We review some in the sequel.

The LRU-Threshold (Chou & DeWitt, 1985) is a simple extension of LRU in which documents larger than a given threshold size are never cached. The LRU-K (O'Neil, O'Neil, & Weikum, 1993) considers the time of the last K references to a

Related Content

Complex Motion Pattern Queries in Spatio-Temporal Databases

Marcos R. Vieira (2016). *Geospatial Research: Concepts, Methodologies, Tools, and Applications (pp. 269-292).*

www.irma-international.org/chapter/complex-motion-pattern-queries-in-spatio-temporal-databases/149497

Artificial Neural Networks

Xiaojun Yang (2009). *Handbook of Research on Geoinformatics (pp. 122-128).*

www.irma-international.org/chapter/artificial-neural-networks/20395

Varying the Variable: Presenting Different Cases for Visualizing the Relative Attractiveness of Retail Business Centers in New Britain, Connecticut

Zachary Klaas (2012). *Geospatial Technologies and Advancing Geographic Decision Making: Issues and Trends (pp. 101-113).*

www.irma-international.org/chapter/varying-variable-presenting-different-cases/63599

GeoCache: A Cache for GML Geographical Data

Lionel Savary, Georges Gardarinand Karine Zeitouni (2009). *Handbook of Research on Geoinformatics (pp. 350-368).*

www.irma-international.org/chapter/geocache-cache-gml-geographical-data/20422

Machine Learning: Definition of Elements and Concepts Dedicated to Bridges

Bertrand Cauvinand Pierre Benning (2017). *International Journal of 3-D Information Modeling (pp. 1-16).*

www.irma-international.org/article/machine-learning/208156