Operation-based revision control of geospatial features

Máté Cserép, Roberto Giachetta

Eötvös Loránd University, Faculty of Informatics

{mcserep, groberto}@inf.elte.hu

Introduction

The evolution and spreading of data capturing methods resulted in an exponential growth in the amount of spatial data maintained by organizations. Usually multiple versions of the same data exist due to application, therefore to manage data efficiently a spatial revision control system is required.

Revision control systems are widely used tools in software development primarily aimed on managing versions of software source code. These systems are generally usable for any text document, however they are not optimized for binary data, as its format is too general for tracking changes within the binary code.

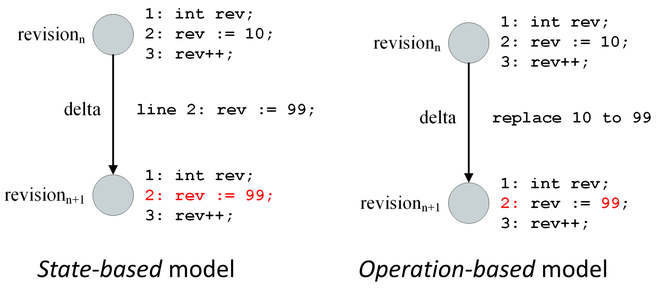
In *GIS* several file formats exist for storing geospatial information, most formats being binary. Although some new solutions have emerged in the revision control of geospatial data (e.g. *GitSpatial*, *GeoGig*), these systems are rather concentrating on specific areas. No global solution has been suggested with respect to all kinds of spatial data and possible operations performed on the data.

Treating spatial information as binary data produces several major drawbacks compared to textual revision control, most notably it erases all semantic information about the applied changes, like the executed spatial operations. Gathering and providing detailed knowledge about any alterations between selected revisions is a particularly important expectation towards a version control system.

Models and methods

There are two main categories for producing the changesets between versions:

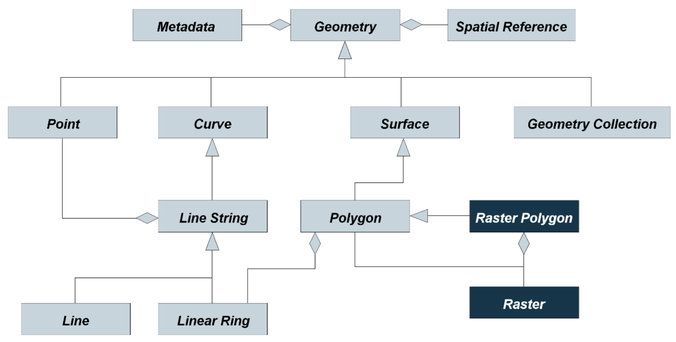
- With *state-based deltas* the general method is to decompose the document into smaller, more easily manageable parts and store the altered state of the changed ones.
- In contrast, *operation-based deltas* store the actually performed operations between two succeeding versions in the revision control system, instead of the alterations in the state. It is notable for this concept to not necessarily requiring any special knowledge about the managed data format from the revision control system to create or merge changesets.



State-based model — *Operation-based* model

Representing spatial data

The data model purely assumes that all geospatial data types in the system directly or indirectly inherit or implement a common ancestor class or interface. A well-known compliant example for the mentioned minimalist specification is the *Simple Feature Access* standard by the *Open Geospatial Consortium*. The SFA is an object-relational mapping for geospatial data, providing an abstract *Geometry* as a common base for spatial objects, including collections.

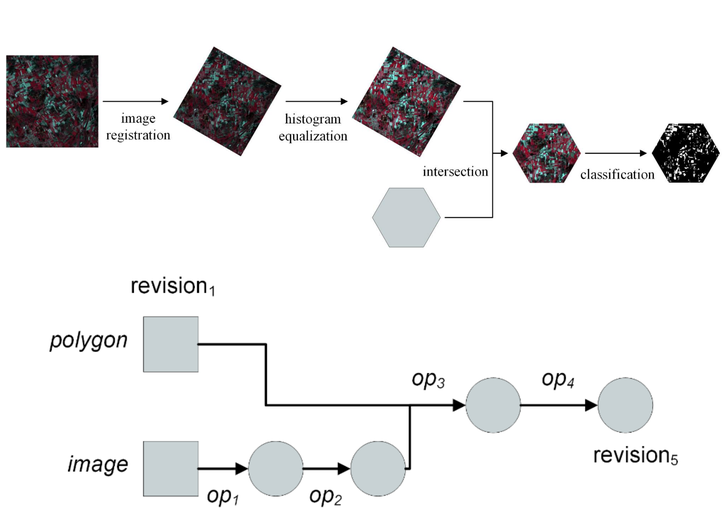


The extended Simple Feature Access data model.

Managing version history

Geospatial operations can be defined as a mapping between geometry objects, and can be represented as transformation objects. The objects can be described by the applied method and the values of arguments (if any). This concept is introduced in the *OGC Spatial Referencing by Coordinates* standard for coordinate transformations, but can be easily generalized for all kind of geospatial operations, including raster image processing. Using this approach, only the descriptors of the operation are required to be stored as operation-deltas, which is compact data, and independent of the size of the actual transformation object.

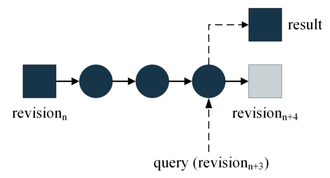
As the core structure for storing the revision history the *directed acyclic revision graph* is used, which was inspired by the idea described in the work of Chen et. al. [1]. In this representation each vertex symbolizes the corresponding version and contains the ordered sequence of operations performed between the current and the predecessor revision, while edges denotes the semantic relationships in the version system. The initial revision contains the source geometries, whilst further revisions are computed using the specified operation.



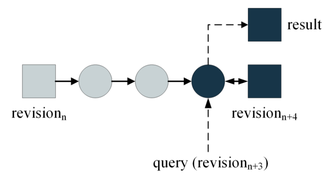
Example workflow with 4 operations and its revision graph.

Version retrieval

In an operation-based revision control system only the initial states of the geometries are needed to be stored, further modifications are represented with transformation objects. In order to balance the efficiency between storage space and revision retrieval execution time, fittingly selected versions can also be snapshoted, which stores the entire geometry set, so it can be retrieved without the necessity of reapplying operations.



Query using *forward deltas*



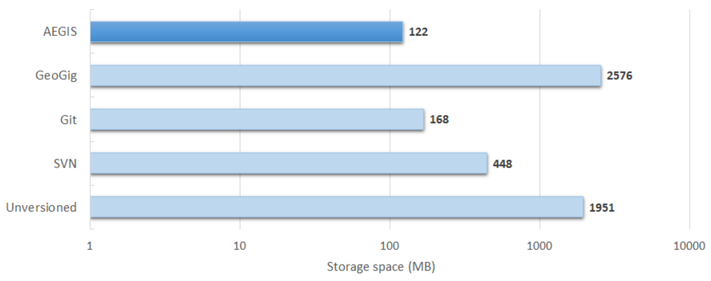
Query using *reverse deltas*

Implementation

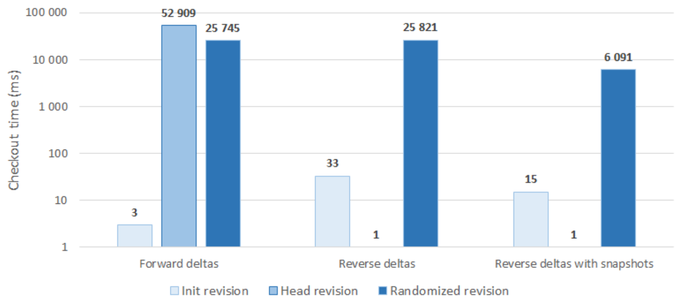
To prove the usability of the presented revision control model the implementation was carried out as part of the *AEGIS geospatial framework* [2] as this system satisfies all requirements described, including a general, abstract data model based on *SFA*, and operation support which utilizes operation metadata. The implementation of both the AEGIS and the revision control system was carried out using the .NET/Mono Frameworks.

Results

The practical applicability of the designed and implemented version control framework can be measured through comparing the created software with similar available tools in the research field. In the interest of quantifying the result of the comparison, storage efficiency and velocity performance, the two most significant attribute of the revision management systems was chosen and examined. In the operation-based model the changesets contain sequences of transformations instead of state modifications, hence the potential drawbacks regarding the computational efficiency must also be thoroughly analysed.



Storage efficiency comparison of revision control tools on the available *OpenStreetMap* dataset of Hungary.



Speed performance comparison of different methods.

References

[1] H.-T. Chen, L.-Y. Wei, C.-F. Chang. Nonlinear revision control for images. In *ACM Trans. Graph.*, 30 (4), pp. 105–115. ACM, New York, NY, USA (2011).

[2] R. Giachetta. AEGIS – A state-of-the-art spatio-temporal framework for education and research. In *OSGeo Journal*, 13, pp. 68–77, OSGeo Foundation, Beaverton, Oregon, USA (2014).