

GENERALIZING XML-ENCODED SPATIAL DATA ON THE WEB

Lassi Lehto and Tiina Kilpeläinen
Finnish Geodetic Institute
Department of Cartography and Geoinformatics
PO Box 15, FIN-02431 Masala, Finland
Fax no: +358-9-29555200
Lassi.Lehto@fgi.fi Tiina.Kilpelainen@fgi.fi

Abstract

The paper presents research and development work concentrating on real-time generalization processes of spatial data for visualization on the Web environment and small devices. The generalisation approach applicable in network-based spatial data services needs to be drastically different compared with the traditional cartographic generalisation. In this paper the use of Extensible Markup Language (XML) techniques in data transformations is presented. The main interest in the research lays in the XSLT process as a means to generalise XML-encoded spatial data in real time. The prototype system developed at the Finnish Geodetic Institute (FGI) applies Geography Markup Language (GML) in spatial data encoding and the data access interface is conformant with the Open GIS Consortium's (OGC) Simple Features specification. The implemented generalization operators include selection, building simplification and aggregation operators, which are run in real-time, during the request-response dialogue in the network.

1. Introduction

The aim of the research presented in this article is the development of real-time generalization of spatial data for visualization in the Web environment and small devices. The starting point for the research is a vision of the future map product as an *ad hoc* representation of the source dataset. As the future map products should be more service-oriented (Kraak and Brown, 2001), the end user of spatial data should have tools to define, by himself, the contents, spatial extent, layout and scale of the final map representation. The resulting temporary map display thus serves as a map in-demand, supporting the various individual needs of a casual map user. The direct access to the most up-to-date geo-databases, available to an individual data user through network, emphasizes the need for sophisticated real-time generalization mechanisms. While the updating methods for the geo-databases are becoming more appropriate (Kilpeläinen 2001), also the casual users expect to access updated geospatial data from anywhere in anytime.

The current map data delivery in the Internet is usually based on raster images. The spatial resolution of the image is decided on the server side, and a proper visualization of the map data is achieved only in that predefined resolution and scale. Some map services provide maps in different scales, but the alternative scales are often based on pre-created raster datasets. If the various scales are produced from a updated database, then all the maps at different scales have to be stored and maintained in the database level beforehand. However, the approach does not allow an arbitrary display scale to be used. A service might support a continuous scale variation, but achieved simply by zooming the original map image without appropriate changes in the displayed map contents or an appropriate generalization level of map objects and symbols. When using the real-time generalization approach the generalized datasets are not stored in the database but rather computed

during the data request, in real time. The use of individual *ad hoc* scales with appropriate content and generalization levels opens up more flexible use scenarios for geospatial databases. The need for management and maintenance of various generalized datasets is reduced once methods for reliable real-time generalization become available.

When using raster image-based map visualization the local interactivity and intelligence of the client application is also rather limited. A map service might provide the end user with a certain level of interactivity, however, usually with high latency and at the expense of a substantial increase in the network traffic. The various new standard developments related to vector-based processing of Web graphics generally, and spatial data in particular, provide promising solutions for development of more sophisticated client applications for visualization (Kähkönen et al. 1999; Lehto et al. 1997).

As a summary, the generalisation approach applicable in network-based spatial data services needs to be drastically different compared with traditional cartographic generalisation. Furthermore, the mobile positioning technologies give an opportunity of having the location of the end user available as an input value for the generalisation process. In the real-time case the resulting map might be generalised in a way that specifically supports the interpretation of environment. In some applications the generalisation might be carried out in a location-dependent way, i.e. things close to the user are presented with higher detail than things farther away. Additionally, in real-time generalisation there is no opportunity to evaluate the result of an individual generalisation process, before it is delivered to the customer. While the traditional generalisation tasks have typically been run as rather time-consuming batch processes, the computation in the real-time generalisation processes must be carried out on-the-fly, during the request-response dialogue in the network. This altogether gives a new challenge for the generalization approach.

2. XML-based Encoding of Spatial Data

Recent developments indicate a drastic change in the mechanisms of Web-based spatial data delivery. The general trend towards XML-based data processing is being recognized also in the spatial data domain (Waters 1999; Zaslavsky et al. 2000). Various standardization communities have been working to develop Extensible Markup Language (XML) vocabularies for encoding spatial data. An interesting example of these is the Open GIS Consortium's Geography Markup Language (GML) recommendation (Lake, 2000). Originally published as OGC's Recommendation Paper in May 2000, the specification is available as an official OGC specification since April 2001 (OGC, 2001). The GML recommendation establishes an XML vocabulary for expressing OGC Simple Features Specification-compliant data in XML syntax.

A basic principle in the design of the XML technology is the total separation of the contents of the data from its presentation characteristics. In the text document processing domain this principle forms a basis for the so-called multi-purpose publishing, an approach in which various presentations, aimed at different end user environments, are produced from a single source (Saarela, 1999). The Extensible Stylesheet Language (XSL) specification is being developed by the W3C as a tool for defining presentation characteristics of an XML dataset (W3C, 2000a). In connection to this work the W3C has created a specification for transforming XML documents, XSL Transformations (XSLT) (W3C, 1999). XSLT is primarily designed for transforming XML documents for presentation purposes. Typical examples include dynamic creation of the table of contents, and creation of a tabular presentation of some data values in the source document. In the graphics domain XSLT could be used, for instance, to transform a dataset from an application-

specific data structure to the new Web vector graphics standard, Scalable Vector Graphics (SVG) (W3C, 2000b).

Several GIS vendors are developing GML support into their software and a few products are already commercially available. It can be assumed that in the near future the GML specification, or some derivative thereof, will become *de facto* standard for spatial data encoding in the Web. Once the format of spatial data content encoding becomes standardized, an opportunity will open for finding a standardized method also for defining visualization characteristics. The most promising technology for this purpose seems to be the XSLT specification, together with the XML-based visualization languages currently under development (Lehto, 2000).

3. XSLT as a Tool for Cartographic Visualization

The XSLT technology provides a powerful tool to define a transformation from the data content encoding language to an appropriate presentation language. The basic functionality that the XSLT mechanism provides is transforming an XML document into another XML document. Therefore the various XML-based visualization languages are the most appropriate form of output from an XSLT process. The languages most interesting for geospatial applications are the Scalable Vector Graphics (SVG), the supposed *de facto* format for Web vector imaging, and the Extensible 3D (X3D) language, the XML-based successor to the popular Virtual Reality Modeling Language (VRML) (Web 3D Consortium, 2001).

The transformation process is depicted in the Figure 1. The input source for the process is provided in the form of an XML dataset as a tree structure and the transformation declarations defined in an XSLT file, designed appropriately for each destination environment. The transformation is carried out by an XSLT Processor software component. The result is again an XML tree structure, expressed in a vocabulary understood by the target device.

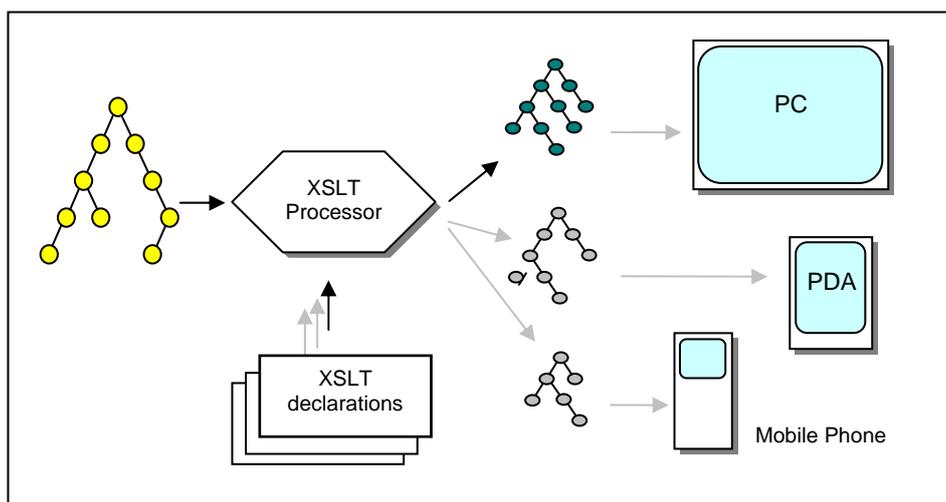


Figure 1. The XSLT transformation process

During the XSLT process the source dataset can be manipulated in various ways. One interesting possibility is to generalize the resulting map image appropriately. So, in addition to translating the dataset into the correct XML vocabulary, an XSLT process can also adapt the map content to the

display characteristics of the end user device used. As the stylesheet technology becomes widely adopted as a visualization mechanism for Web-based spatial applications, it will, for the first time in the history of GIS, provide a standardized means for defining map symbology.

4. A Prototype System and Examples for Real-Time Generalization

A working prototype system has been developed at the Finnish Geodetic Institute (FGI) for testing the above-mentioned techniques (Lehto and Kilpeläinen, 2000). The prototype is based on a three-tier processing model. The first level consists of a Smallworld GIS database server. The middle tier has been developed in-house as a Java Servlet-based Web server extension. Communication between the Smallworld GIS and the middle tier is based on the CORBA technology and on an OGC Simple Features (SF) specification-compliant access interface provided by Smallworld. This is going to be replaced by the SIAS server, a product of GE Smallworld. OGC's GML is being used as the data encoding language. A client application on the third level has been built using a free map-visualization Java library, called OpenMap (BBN Technologies, 2000). Another client application tested is the SVG Viewer browser plugin from Adobe (Adobe, 2001).

Major part of the Java development of the prototype done at the FGI is related to the communication between the middle tier and the Smallworld Simple Features CORBA server and to the construction of the XML source tree from the received data. The source tree is constructed according to a proprietary data model, designed to facilitate easy XSLT transformations to the known destination data models.

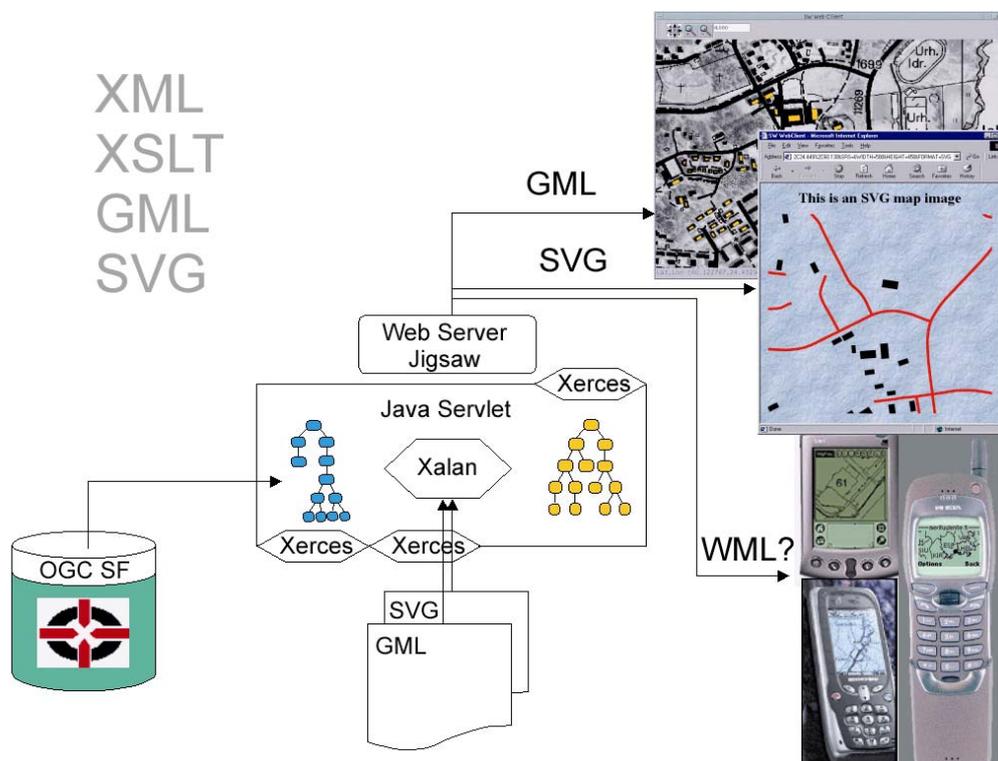


Figure 2. Real-time generalization of XML-based GI

The prototype architecture is presented in the Figure 2. The free-of-charge XML software components applied in the system include an XML parser (Xerces) and an XSLT processor (Xalan) from Apache community (Apache, 2001). The client examples shown in the figure are OpenMap-based Java client application with a GML supporting layer developed at FGI, and the Internet Explorer Web browser with Adobe's SVG Viewer plugin. The future development is to concentrate on mobile platforms and their data encodings like Wireless Markup Language (WML) or Extensible HTML (XHTML).

The XSLT files developed at the FGI provide an example of the results that can be achieved when transforming spatial data by an XSLT processor. The XSLT specification is a promising tool as a solution to the need of generalizing spatial datasets in real-time. Most simple generalization operations, like filtering out the unneeded parts of the dataset, and selecting from among alternative geometries, are readily available. More sophisticated generalization tools can be added via the XSLT extension mechanism. Typical examples include coordinate manipulations, like line smoothing. The generalized datasets are written out as XML data and can thus be easily visualized in various XML-conscious client applications. In the research project the operators selection, simplification of building outlines and aggregation of building symbols, have so far been implemented by using the XSLT mechanism.

The emphasis of the intellectual work in building an XSLT-based service naturally concentrates on the design of appropriate XSLT files. The XPath expressions in the templates pick up the desired part of the source data for processing. The contents of the templates define the structure of the result tree. The XSLT files developed at the FGI provide a demonstrative example of the generalization results that can be achieved when transforming spatial data by an XSLT processor (see Figure 3).

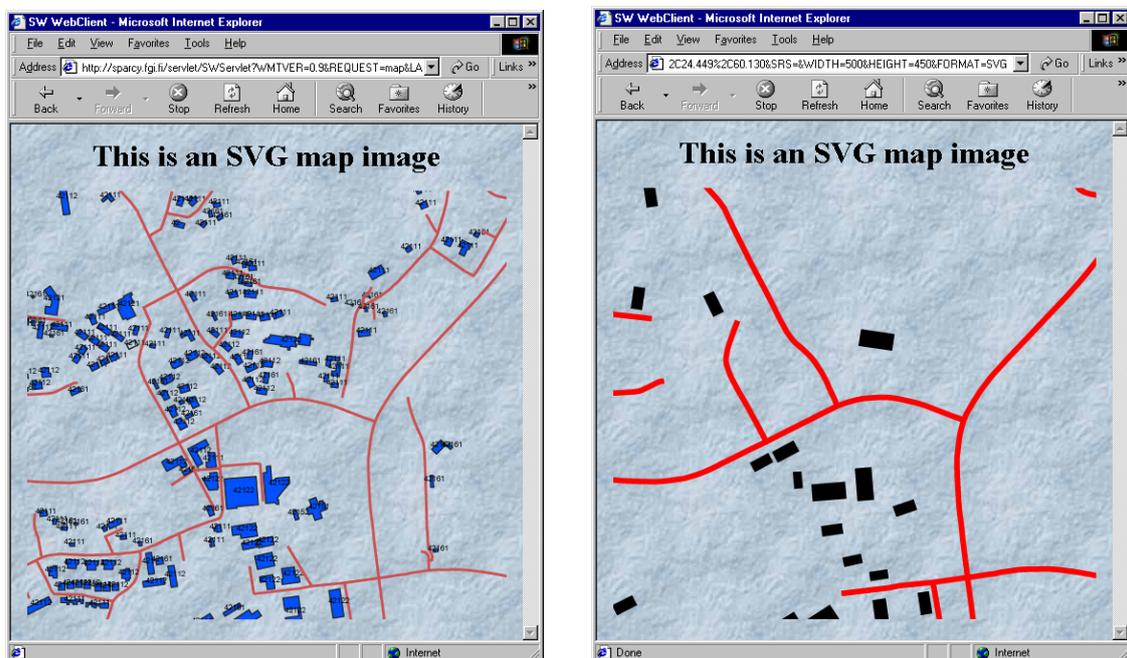


Figure 3. Two SVG displays of the same dataset, transformed by different XSLT processes. Two generalization methods have been implemented and tested: selection and simplification.

In the implemented selection function individual spatial objects can be selected or rejected for inclusion in the result dataset based on their feature type. Decision can also be based on computations performed during the transformation. The extension mechanism available in the XSLT process enables arbitrary, application-specific functions to be introduced into the transformation process. Several XSLT processes can also be chained together, if the task is too complicated to be expressed as one individual transformation.

5. Concluding remarks

The first experiences of the prototype implementing real-time generalization of XML-encoded spatial data are encouraging. The use of freely available components, like the XSLT processor and the XML parser, facilitate the programming work considerably. Although the system is based on a nascent technology, most of the available components are fully functional and reliable. Poor performance is clearly an issue. This is partly due to the inefficient character encoding applied in XML, partly to the poorly optimized beta category software. The focus of the research in the FGI has been on the generalization aspects of the XSLT process. In this respect a considerable progress has been achieved by developing a set of Java-based XSLT extension functions that carry out the most complicated processing tasks needed in the generalization process. The work continues with a special attention being paid on the issues related to the use of maps in mobile devices.

References

- Adobe, 2001. "The SVG Zone", www.adobe.com/svg/index.html
- Apache, 2001. The Web site of Apache XML Project, xml.apache.org
- BBN Technologies, 2000. "OpenMap, Open Systems Mapping Technology", www.openmap.org
- Gould, M. and A. Ribalaygua, 1999. A New Breed of Web-Enabled Graphics", *GeoWorld*, March 1999, pp. 46-48.
- Kilpeläinen, T., 2001. Maintenance of Multiple Representation Databases for Topographic Data. *The Cartographic Journal*, Volume 37 Number 2: 101–107.
- Kraak, M-J. and A. Brown, 2001. *Web Cartography, developments and prospects*. Taylor & Francis Inc, London, 209 p.
- Kähkönen, J., Lehto, L., Kilpeläinen T. and T. Sarjakoski, 1999. Interactive Visualisation of Geographical Objects on the Internet. *International Journal of Geographical Information Science*, Taylor & Francis Ltd, Volume 13, Number 4, pp. 429-438.
- Lake R., 2000. "Geography Markup Language – The Future of Internet Mapping", www.web-mapper.com/articles/GML.html.
- Lehto, L., J. Kähkönen and T. Kilpeläinen, 1997. WWW-Technology as Means of Transfer and Visualization of Geographic Objects. *Proceedings of the 18th ICA/ACI International Cartographic Conference*, Stockholm, Sweden, June 23–27, 1997, Vol. 2, pp. 681–689.
- Lehto, L., 2000. XML in Web-based Geospatial Applications. *3rd AGILE Conference on Geographic Information Science*, Hanasaari Conference Centre, Helsinki/ Espoo, May 25–27, 2000, pp. 162–167.
- Lehto, L. and T. Kilpeläinen, 2000. Real-time Generalization of Geodata in the Web. *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, the Netherlands, 16–23 July 2000, Vol. XXXIII, Part B4, pp. 559–566.
- OGC, 2001. "Geography Markup Language (GML) 2.0" www.opengis.net/gml/01-029/GML2.html

- Saarela, 1999. "The Role of Metadata in Electronic Publishing", Doctoral Dissertation, Helsinki University of Technology, Espoo, 1999.
- W3C, 1999. "XSL Transformations (XSLT) Version 1.0", www.w3.org/TR/xslt.html
- W3C, 2000a. "Extensible Stylesheet Language (XSL) Version 1.0", www.w3.org/TR/xsl/index.html
- W3C, 2000b. "Scalable Vector Graphics (SVG) 1.0 Specification", www.w3.org/TR/2000/CR-SVG-20001102/index.html
- Waters, N., 1999. Is XML the Answer for Internet-based GIS?, GeoWorld, June 1999.
- Web 3D Consortium, 2001. "Extensible 3D (X3D) Task Group", www.web3d.org/TaskGroups/x3d/index.html
- Zaslavsky I., Marciano R., Gupta A. and Baru C., 2000. XML-based Spatial Data Mediation Infrastructure for Global Interoperability. 4th Global Spatial Data Infrastructure Conference 13-15 March 2000, Cape Town, South Africa.