

Gathering statistics of XML Schema Usage in OGC Web Services

Alain Tamayo, Carlos Abargues, Arturo Beltrán, Carlos Granell, Joaquin Huerta
Institute of New Imaging Technologies, Universitat Jaume I, Spain

Keywords: OGC Web Services, XML Schema, Complexity analysis

Introduction

OGC Web Services (OWS) are a set of service interface specifications to exchange geospatial information. These specifications use XML Schema [1][2] as the language of choice to express the structure of the exchanged messages. The size of the schemas associated to the specifications has been growing with time reaching a point that they are very hard to understand and use to build real systems. Studying the complexity of these specifications and finding ways to simplify them is a topic that can help in reducing systems development time. One important aspect needed to study the schemas complexity is gathering statistics of schema features usage through the different specifications.

In this paper we present a module that allows users to gather and visualize statistics about XML Schema usage in OWS. The module analyses schema files associated to the specifications and records how often the different schema features (types, elements, attributes, etc.) are used through the specifications. These figures give us an idea of how big and complex the different specifications are. It also provides several options to visualize, combine, and compare the information gathered from different specifications.

The tool has been implemented as an add-on to OGC Schemas Browser [3], an open source tool focused on the visualization of dependency relationships between different entities on the specification schemas. This browser provides support for the visualization of specification dependencies, namespace dependencies, file dependencies, and type dependencies. It also checks for errors that might go undetected when using XML editors that do not perform validation in a thorough way.

XML Schema Metrics

XML schemas metrics are mostly based on metrics build to measure different parameters on software systems or on XML documents [4][5][6]. In [7] and [8] a comprehensive set of metrics for measuring schemas complexity are presented. These metrics are the starting point of our work. The use of metrics allows us to quantify the complexity, quality and other properties of the schemas. This way we can have a finer control of the design and evolution process for the schemas, ultimately improving the quality of the final products based on them. The metrics calculated by our module for the OGC Web Services schemas are listed next:

- *Lines of Code*
- *Number of files*
- *Types:*

- Total number of complex types
 - Number of complex types defined with sequence
 - Number of complex types defined with choice
 - Number of complex type defined with all
- Total number of simple types
 - Number of simple types defined as lists
 - Number of simple types defined as unions
- Number of anonymous types
- *Number of global elements*
- *Number of global model groups*
- *Number of global attributes*
- *Number of global attribute groups*
- *Number of global items (types, elements, model groups, attributes and attribute groups)*
- *Number of wildcards*
- *Number of abstract types and elements*
- *Number of substitution groups*
- *Number of times derivation by extension is used*
- *Number of times derivation by restriction is used*

OGC Schemas Browser

OGC Schemas Browser (Figure 1) is an open source tool designed to browse OWS specifications schemas in an intuitive way that complement other XML editors such as *XMLSpy*¹ and *oXygen*/². This tool presents different views that help understanding the relationships between files, specifications and types. Some of the features it offers are:

- Visualization of specification and namespace dependencies
- Visualization of file dependencies
- Visualization of type dependencies
- Detection of XML Schema design errors

The core of the application is the *schema processor*, which is in charge of processing the schema files and all of their dependencies, building an internal representation that will be used later to generate graphs of elements relationships and to calculate the software metrics mentioned in the previous section. The browser uses the libraries *JGraph*³ to display graphs of entity relationships and *Eclipse XSD*⁴ to parse XML schemas.

Displaying Statistics of XML Schema Usage

OGC Schema Browser allows the processing and visualization of one specification at a time. When the specification is loaded statistics of XML usage is gathered and persisted in a repository. Metric values for a given specification are displayed by default in the *Output* section of the browser when at loading time.

¹ <http://www.altova.com/xml-editor>

² <http://www.oxygenxml.com/>

³ <http://www.jgraph.com/>

⁴ <http://www.eclipse.org/modeling/mdt/?project=xsd#xsd>

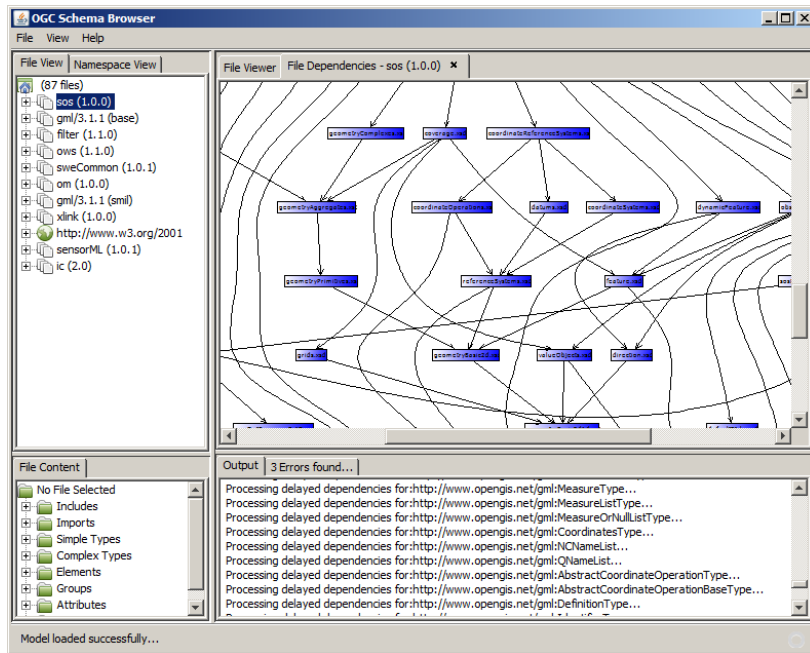


Figure 1: OGC Schemas Browser user interface.

Sometimes it is useful to compare statistics of one specification with those of other specifications. To do this we can use the features offered by the *visualization module*. This module provides several options to display, combine, and compare the information found on the repository of XML Schema usage statistics. This module is based on *JFreeChart*⁵, a Java library designed to display charts. *JFreeChart* allows the visualization of several kinds of graphs. In our case we will use pie charts, bar charts, and single and dual axis plot charts.

Pie charts are used to show the proportions of the different schema components per specification. For example, Figure 2 shows how global schema components for the SOS 1.0.0 specification [9] are distributed.

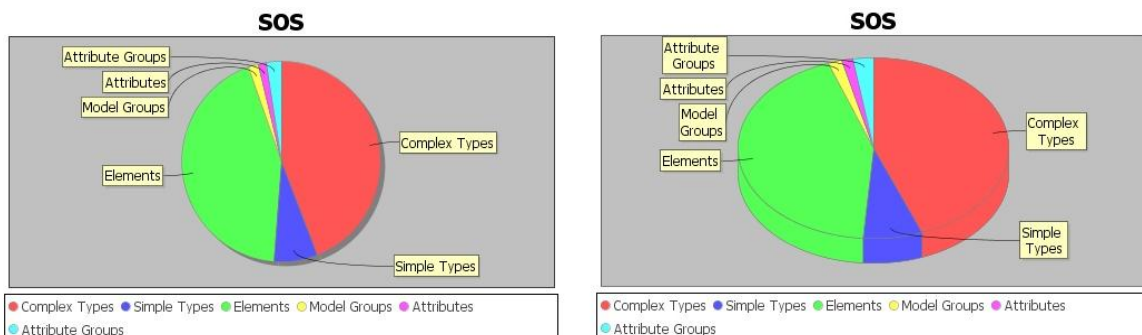


Figure 2: Pie chart showing the decomposition of SOS in schema component types

Global schema components are divided into complex and simple types, elements, attributes and elements and attribute groups. Pie charts can be displayed in 2D or 3D versions. The same information can be also displayed using bar charts (Figure 3).

⁵ <http://www.jfree.org/jfreechart/>

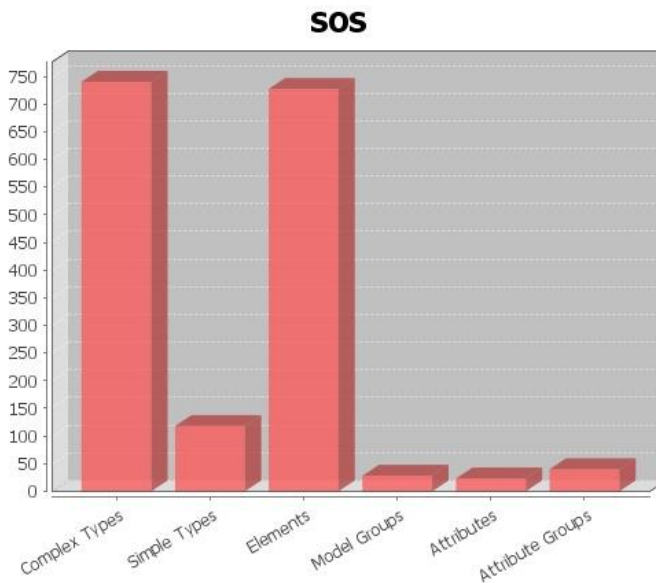


Figure 3: Bar chart showing the decomposition of SOS in schema component types

Bar charts are also used to compare metrics between different specifications. For example Figure 4 depicts a bar chart that compares the number of global schema components between SOS 1.0.0 and WFS 1.1.0 [10]. The users may select which metrics from which specifications they want to compare. This kind of charts can also be shown in 2D and 3D versions.

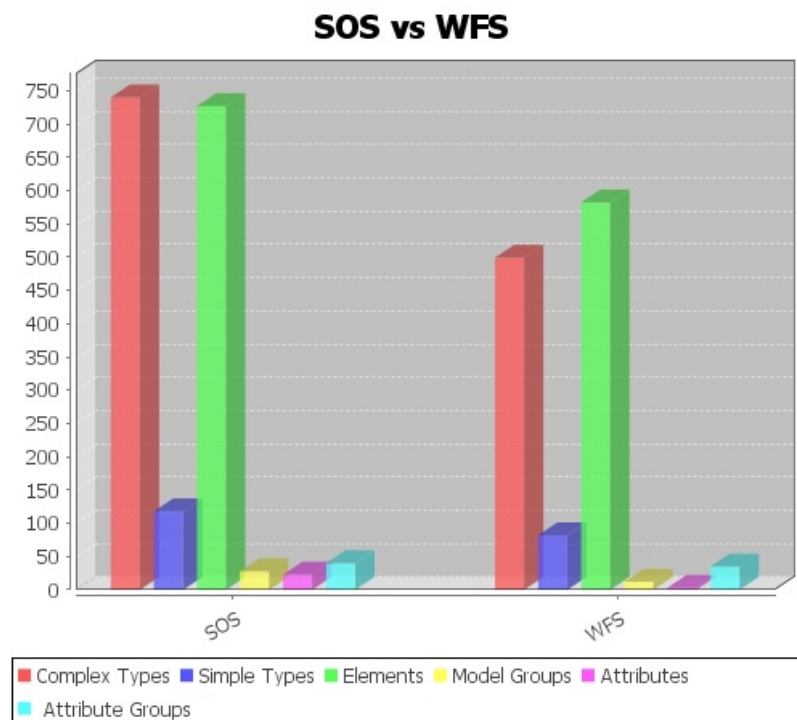


Figure 4: Comparing metrics between different specifications

Last, users may show values from two different metrics from a set of specifications in a plot diagram. This may be useful to detect possible correlation relationships between two different metrics. For instance, Figure 5 shows the comparison between the number of complex types and the number of elements throughout the specification schemas.

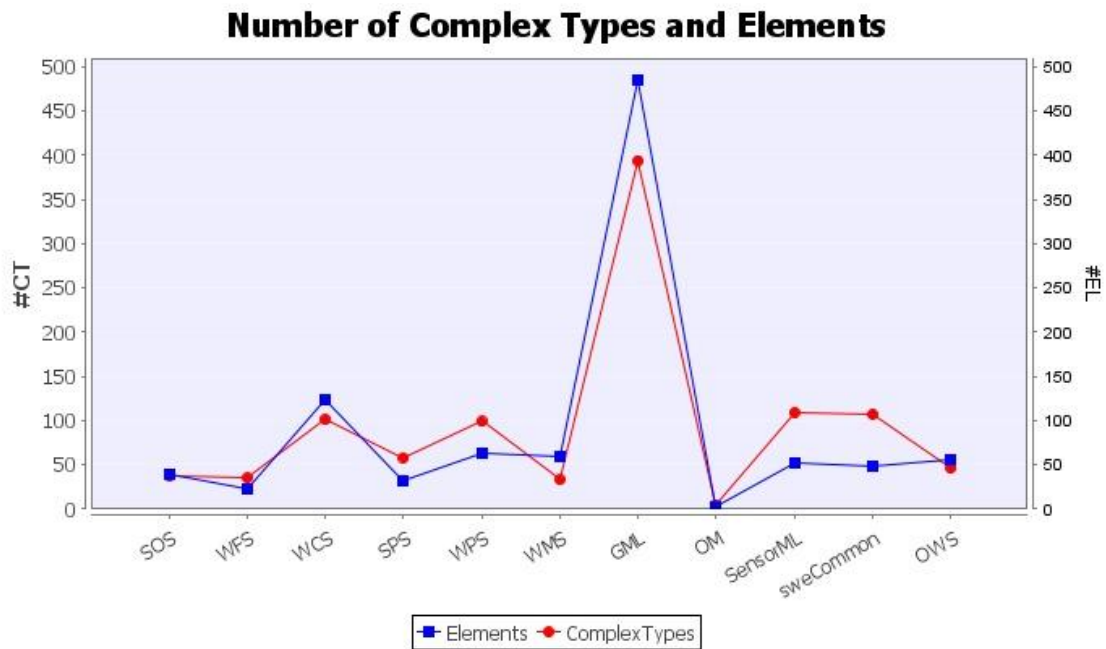


Figure 5: Dual axis plot graphs

Conclusions

In this paper we have presented a module that allows users to gather and visualize statistics about XML Schema usage in OWS. It analyses schema files associated to the specifications and records how often the different schema features (types, elements, attributes, etc.) are used through the specifications. Gathering these values allows us to quantify the complexity, quality and other properties of the schemas. This way we can have a finer control of the design and evolution process for the schemas, ultimately improving the quality of the final products based on them. The tool has been implemented as an add-on to OGC Schemas Browser, and allows the visualizations of pie, bar and plot charts, combining the information gathered from several specifications.

References

1. W3C: XML Schema Part 1: Structures Second Ed., <http://www.w3.org/TR/xmlschema-1>
2. W3C: XML Schema Part 2: Datatypes Second Ed., <http://www.w3.org/TR/xmlschema-2>
3. Tamayo A., Granell C., Huerta J. OGC Schemas Browser: Visualising OWS XML Schemas. Accepted short paper AGILE 2010, Guimaraes, Portugal, May 2010.
4. Barbosa D., Mignet L., Veltri P.: Studying the xml web: Gathering statistics from an xml sample. *World Wide Web*, 9(2):187-212 (2006)
5. McCabe T.: A Complexity Measure, *IEEE Transactions on Software Engineering*, vol. Se-2, no. 4, pp. 308-320 (1976)
6. Qureshi, M.H.; Samadzadeh, M.H.: Determining the complexity of XML documents. *Int'l Conference on Information Technology: Coding and Computing*, 416 - 421 Vol. 2 (2005)
7. Lammel R., Kitsis s., Remy D.: Analysis of XML schema usage. In *Proceedings of XML Conference*, pp. 1-35 (2005)

8. McDowell A.,Schmidt C., Yue K.: Analysis and metrics of XML schema. In Proceedings of Intl Conference on Software Engineering Research and Practice,pp. 538-544 (2004)
9. OGC: Sensor Observation Service. 1.0.0. OGC Doc. Number 06-009r6 (2007)
10. OGC. OpenGIS® Web Feature Service Implementation Specification. Version. 1.1.0. OGC Doc. Number 04-094 (2005)