Assertion Support for Manipulating Constrained Data-Centric XML

Patrick Michel

Software Technology Group

University of Kaiserslautern

14.04.2009

XML is used for different purposes. We are interested in data-centric applications of XML where it is used to handle structured data in loosely coupled, distributed systems. In many such scenarios, it is important that the XML data complies to structural and integrity constraints, in particular to value-based constraints. The constraints should remain invariant under operations manipulating the data.

In [1], we presented a technique to define and maintain invariants. For a core procedural manipulation language, we showed how to automatically derive the weakest precondition of procedures for the constraints. We presented a path-based formalization of a data-centric XML abstraction and an assertion language that enables weakest precondition generation.

By focusing on paths it is possible to significantly simplify preconditions. This is done by isolating the aspects that are manipulated by the considered procedure from the aspects that remain unchanged. As the overall invariant can be assumed for the input data, only the isolated aspects have to be checked in prestates of procedures.

Building on these theoretical foundations, we are developing higher-level specification languages, which are much more similar to widely used ones, yet can be completely translated into the present formalization. For structural constraints we propose a grammar-based approach in the form of a dialect of Relax NG. For integrity and value-based constraints, we propose a rule-based approach using a specification language resembling XPath. This language has a liberal syntax and is embedded as annotations into the structural grammar.

Although such combinations have been proposed before, we are able to define the semantics of the resulting specification within the path-based formalization and are so able to automatically generate and simplify weakest preconditions for local manipulations. The combination of structural patterns and XPath-like constraints allows to dramatically shorten specifications, providing a very convenient syntax usable by domain experts.

Building on such a specification, the domain expert can write local, atomic manipulation procedures, which represent domain-specific tools necessary to evolve the represented data. Again, we are offering a convenient syntax to define such procedures, which are then translated into the core language presented in [1].

All these techniques are targeted at practical application, ultimately an integration into a general purpose programming language. For this reason traceability of constraints and adequate error messages are necessary to be able to react properly to the failure of a precondition. Every precondition can therefore be traced back to a specific source location, is categorized and has access to concrete values and paths which contribute to it. This information can be used on failure to create a proper response and eventually even proceed by invoking other manipulation procedures.

In this talk I will give a short overview of the general approach and demonstrate our prototype system with practical examples.

References

[1] Patrick Michel and Arnd Poetzsch-Heffter. Assertion support for manipulating constrained data-centric xml. In *International Workshop on Programming Language Techniques for XML (PLAN-X 2009)*, January 2009.