

Abstract

# Converting XML/DTD to Parsing Expression Grammars

SHINYA YAMAGUCHI<sup>1,a)</sup> KIMIO KURAMITSU<sup>1</sup>

Presented: March 10, 2015

A Document Type Definition is a grammar for describing the structure of an XML document. Today, XML is a standard format to exchange data on the Web, and the validation of exchanged data is crucial. DTD is the most common schema among many XML validation languages such as XML Schema and Relax NG, which requires an additional validator implementation to an XML parser. On the other hand, Parsing Expression Grammars is a popular foundation for describing syntax. In general, PEGs are more expressive than CFGs. In this paper, we address the conversion from XML/DTDs to PEGs. We present that all DTDs can be mapped to PEGs, which implies that PEGs contains the expressiveness of XML/DTDs. Due to the conversion, arbitrary PEG-based parser can validate XML documents with converted DTDs. We will show that our implemented parser achieves a competitive performance to existing XML/DTD validators even if the input size is extremely large.

---

<sup>1</sup> Yokohama National University

<sup>a)</sup> y.shinya.kml@gmail.com