

C++ Implementation of a Two Phase XML Validator

Jennifer Tom

Akimeka LLC

Research Supervisor: Christopher Paris

Home Institution: Maui Community College

XML schemas are used to validate the structure of XML documents by defining the content and structure of the XML document. A two phase XML validator was implemented using C++. In phase one, a state machine is created from an XML schema. In phase two, the XML documents are validated using the state machine from phase one. Separate programs were written for each phase to keep the validator lightweight and to make validation faster. The validator handles schemas written in a subset of the Relax NG schema definition language. Phase one of the validator reads a schema using a parser built from the EXPAT XML parser toolkit. This parser creates an object model of the schema. A compiler was then written to convert the object model into a state machine. Phase two consists of a driver program to execute the state machine on an input XML document. The parser was tested using a test XML schema based on the Relax NG subset. My validator will be compared to a mature validator based on a variety of schemas and XML documents.

Jennifer Tom is currently attending Maui Community College, taking classes to fulfill prerequisites to apply for the Master's in Information and Computer Science program at UH Manoa. She has a BS in Biochemistry from Washington State University and is looking to enter the field of Bioinformatics. Specifically, Jennifer would like to design programs for researchers in the life sciences. Her dream is to write programs for researchers in the life sciences to help solve complex problems, such as predicting protein folding!

