

Application of XML for future EXFOR software

M. Chiba

Nuclear Data Exchange Format of Next generation

- I suppose: Nuclear Data Exchange Format of Next Generation should be defined based on XML
- Current EXFOR or NRDF Format is specific to itself
- If the Exchange Formats be redefined based on XML, we might utilize various XML supports available in developing applications

XML

- XML document: text content marked up with text tags
- Well-formed XML document: XML parsers can read it and understand it
- Tag Syntax:
 - start tag: <name of element>
 - End tag: </name of element>
- XML tree
- Attribute

DTDs and schema

- XML documents can be validated by XML parsers with DTDs or schema
- XML documents containing Well-formedness mistakes can be checked before publishing it

Validating parser

- Validating parser compares a document to its DTD and lists any places where the document differs from the constraints specified in the DTD
- Validating a Document:
 - Use the parser's API to validate documents
 - Use one of the online validators
 - Use a local program to validate the document

Namespaces

- Purposes
 - To distinguish between elements and attributes from different vocabularies with different meanings that happen to share the same name
 - To group all related elements and attributes from a single XML application together so that software can easily recognize them

Internationalization

- Character Set
- Encoding
- Text declation

XML as Document Format

- XML: first and foremost a document format
 - Human beings would read
 - A syntax for computer data in applications
 - Rigid structures: tree
 - Narrative Document Structures

XML on the Web

- XSLT stylesheet

XSL Transformations(XSLT)

- `xsl:stylesheet` and `xsl:transform`

XPath

- XPath: a non-XML language for identifying particular parts of XML documents
 - XSLT uses Xpath expressions to match and select particular elements in the input document for copying into the output document or further processing
 - Xpointer uses XPath expressions to identify the particular point in or part of an XML document to which an XLink links
 - W3CXML Schema Language uses XPath expressions to define uniqueness and identify constraints
 - XForms relies on XPath

XLinks

- Xlink: an attribute-based syntax for attaching links to XML documents

XPointers

- XPointer: a non-XML syntax for identifying locations inside XML documents

XInclude

- Xinclude: a new technology for combining multiple well formed and optionally valid documents and fragments thereof into a single document

Transformation from a Semantic markup into a presentational markup

- Cascading Style Sheets(CSS)
- XSL Formatting Objects(XSL-FO)

XSL Formatting Objects

- XSL-FO:

XML as a Data Format

- Inventors intention: XML as a format for narrative documents to be read by people
- Most common applications of XML today: storage and transmission of information
- The structures appropriate for applications
 - Rigid
 - Strongly typed element
 - Elements tend to look more likely database records
 - Make data portable
 - Proliferation of free XML tools

Strengths of using XML

- Strengths of using XML as a software data format:
 - Simple syntax: easy to generate and parse
 - Support for nesting: Nested elements allow programs to represent complex structures easily
 - Easy to debug: human-readable data format is easy to explore and create with a basic text editor
 - Language- and platform-independent: XML and Unicode guarantee that your data files will be portable across virtually every popular computer architecture and language combination in use today

XML Schemas

- Many applications need a powerful and expressive validation method:
 - W3C developed the XML Schema Recommendation
- Schemas can describe complex restrictions on elements and attribute
 - An XML Schema is an XML document containing a formal description of what comprises a valid XML document

Schemas Versus DTD

- DTD provides the capability to do basic validation of the following items in XML documents:
 - Element nesting
 - Element occurrence constraints
 - Permitted attributes
 - Attribute types and default values
- Not provide fine control over the format and data types of element and attribute values

Schemas Versus DTD

- W3C XML Schema standard includes the following features:
 - Simple and complex data types
 - Type derivation and inheritance
 - Element occurrence constraints
 - Namespace-aware element and attribute declarations

Programming Models

- XML support is available for virtually every programming platform in use today
 - Text-based XML processing
 - Event-driven XML processing
 - Tree-based XML processing
 - Pull-based XML processing

Document Object Model(DOM)

- DOM: an API for accessing and manipulating XML documents as tree structures

Simple API for XML(SAX)

- SAX: event-based API for reading XML documents