



MIE 2005 Tutorial #300

Title: Introduction into XML Standards

List of instructors: Prof. Joachim Düdeck

Date, time and place: Sunday, 28 August 2005, from 1pm to 4pm, room 2160

Abstract: XML is a metalanguage for structuring and storing textual documents. The main purpose is the system independent processing of documents by separating content, structure and presentation. Several hundred XML based languages have already been defined and are in use in e-business, e-health and other domains. The Tutorial will introduce into the main concepts of XML itself and in the most important XML standard languages (XML Schema, Namespaces, XPath, XSLT, Stylesheets, XForms, XQuery, RDF, Topic Maps etc.).

The XML representation of HL7 messages and the Clinical Document Architecture CDA as XML languages in the medical domain will be explained. Knowledge representation with Topic Maps and its application in ICD10, SNOMED CT etc. are finally described.

The participants will get an impression of the power and applicability of the XML concepts and its current use in particular in the health care environment

Description: XML is going to become the Web standard for representing structured textual information in different types of documents. It has been designed by the Worldwide Web Committee (W3C) which is also responsible for the continuous maintenance and further developments. XML is a metalanguage which is the basis of meanwhile about 50 standard languages (XML Schema, XSLT, XSL Stylesheets, XForms etc.) also designed by the W3C and of an additional set of several hundred application languages in quite different domains (CDA, SOAP, WSDL etc.). It is also the basis of the Semantic Web development. For the successful application of these standards a fundamental understanding of the available XML languages and their potentialities and use is required.

The XML syntax can be understood quite easily. XML documents consist of Elements, Attributes and tags. The content of elements is included into opening and closing tags which start with < and end with >. Closing tags start with </. Besides the / the content of opening and closing tags have to be case-sensitive identical. Elements can be hierarchically nested, by including start and opening tag of elements within the tags of another element. The XML document can be understood as a tree with element nodes on different hierarchical levels. Each XML document is embraced by the document node.

Attributes are mainly used for representing meta-information about the element content (e.g. attribute "codesystem" as meta-information of the "code value"). Other applications are identifier attributes, linking attributes etc.. Attributes are always presented in the opening tags of elements.

Content and structure of XML documents can be described by Document Type Definitions (DTD) or XML Schema. In DTD's a non-XML language is applied. DTD's are easy to understand, but limited in its functionality. They will be replaced more and more by XML Schema. Schemas are very powerful XML-based documents for the detailed definition of content and structure of XML documents. The XML Schema definition provides a large number of predefined, built-in datatypes. The content of elements and attributes is defined by using simpleType elements, the structure by applying complexType elements. Schemas can be composed from different schema fragments.

Schema support namespaces to avoid mismatches in particular of element tags defined in different environments or Schema fragments. Namespaces use qualified names with namespace prefixes (<lab:diagnose> - diagnose tag as defined in the lab namespace). This approach facilitates the processing of XML documents and the composition of schemas from different fragments.

XML documents which fulfil the requirements of a DTD or Schema are called "valid". Other XML documents are only "wellformed".

XML documents are in general transparent to the user. They are the container in which the content is system independently stored. XML documents are generated by user interfaces which can be constructed from XML document instances or from XML Schema using XForms as an XML language or other tools like Infopath from Microsoft.

For the presentation of XML documents stylesheets are used. The presentation process can be decomposed into three distinct steps, the navigation with XPath, the transformation into the new document with XSLT and the presentation using HTML or the more comprehensive stylesheet language XSL-FO.

With XPath different nodes in the original document, the "source tree" can be addressed and transformed with XSLT into the structure of the new document, the "result tree". This document can consist of plain text, it can be structured e.g. as an HL7 message or can be another XML document even with newly defined elements and attributes.

The most important advantage of an XML environment is the retrieval capacity. Using XQuery very specific content of the XML documents can be retrieved. XQuery uses also XPath for navigating in the XML documents.

Besides the already described Standard Languages a more and more increasing number of XML application languages have been defined. In Medicine the Clinical Document Architecture CDA is becoming a widely used standard for defining and storing clinical documents as a first step towards an Electronic health care record. Also HL7 in version 2.x, but mainly in the newly defined version 3 is using XML as interchange format. These XML applications will be explained in detail at the end of the tutorial.

The participants will get a comprehensive overview on the power of storing and processing concepts of documents and messages based on XML.

Who should attend ? Medical professionals and IT specialists who are interested resp. have to deal with XML technology in their environment.

Prerequisites: No XML experience is required.