Summary

The course is about the foundations and tools for processing tree structured data, a prevalent model for representing semi-structured data (SSD) over distributed information networks. It aims at presenting approaches, programming languages and tools for modeling and manipulating tree-structured info.

Content

The theoretical part introduces underlying concepts sustaining the approach.

The practical part illustrates the application of the concepts in a concrete context: the development of Web applications that make use of an XML native database (one category of the NoSQL databases) and associated XML languages.

Theoretical foundations

- Tree grammars
- Finite tree automata
- Type systems to describe and validate the structure of SSD
- Document Type Definition
- XML Schema
- RELAX NG and Schematron
- Querying tree structured data and programming
  - Navigation and extraction of information from tree structured data (XPath expressions)
  - Tree data transformation (XSLT)
  - Query and programming language (XQuery) incl. Static Type Checking

Application scenario

- Use of a development framework in which all these languages fit

Keywords

Tree-shaped data representation and processing, Foundation of XML types, Tree grammars, XML core technologies, Web applications

Learning Outcomes

By the end of the course, the student must be able to:

- Explain and understand the differences - strengths and weaknesses - of a tree structured model in comparison with other data models.
- Understand the fundamental principles of a strongly typed language to manipulate tree structured data.
- Use core languages for modeling, querying, repurposing and processing tree structured data.
- Identify situations where information management requirements can be more appropriately dealt with a tree structured data model approach.
- Get a flavor of research ongoing in the domain.

Teaching methods
Ex cathedra lectures and group mini-projects.

**Expected student activities**
- Attend the lectures
- Work on mini-project

**Assessment methods**
- Written exam and mini-project evaluation.