

Course: Foundations for Processing Tree-Shaped Data with Applications to XML Technologies

Pierre Genevès
pierre.geneves@cnrs.fr

November 22, 2016

This course aims at introducing principles and foundations for the correct and efficient processing of tree-shaped data. It is illustrated in the setting of XML and NoSQL technologies. The course roughly consists of three parts: an introduction to the most essential XML technologies; an introduction to the relevant theoretical tools that guide the development of these technologies; and an introduction to recent research in the area. The theoretical part introduces tree grammars, finite tree automata, classical tree logics. Then, advanced static analyses are introduced through a μ -calculus of finite trees. The course explains why those tools are important and how they are applied to practical problems, including scalable validation of document streams, efficient query evaluation, static analysis of complex queries in the presence of schemas, static type checking of programs manipulating structured data and documents. The course also aims at presenting challenges, important results, and open issues in the area of NoSQL programming.

Course outline

1. Part 1: Technologies

- Essential XML technologies: Core XML, Schema Languages, Parsing
- Streaming Validation (with SAX and DTD)
- Queries and transformations (XPath, XQuery)

2. Part 2: Fundamentals

- Foundations of XML Types (Tree Grammars, Finite Tree Automata)
- Tree Logics (First-Order Logic, Monadic Second-Order Logic)
- Tree Logics continued (μ -calculus)

3. Part 3: Recent Research

- Query Analysis
- Static Type Checking for XQuery and NoSQL languages
- Grand Challenges and Current Research