

Leonid Libkin - Incomplete data management

Semistructured data is an umbrella term encompassing data models which are not logically organized in tables (i.e., the relational data model) but rather in hierarchical structures using markers such as tags to separate semantic elements and data fields in a 'self-describing' way. In this lecture we survey some of the multiple connections between formal language theory and semi-structured data, in particular concerning the XML format. We will cover ranked and unranked tree automata, and its connections to Monadic Second Order logic, First Order logic, and XPath. The aim is to take a glimpse at the landscape of closure properties, algorithms and expressiveness results for these formalisms.