

# *XML Document Processing: Theory and Practice*

Course note based on

*Taxonomy of XML Schema Languages Using Formal Language Theory*, ACM Transaction on Internet Technology, 5(4):660-704,  
authored by Makoto Murata, Dongwon Lee, Murali Mani, and  
Kohsuke Kawaguchi

course note prepared by  
Tyng-Ruey Chuang

Institute of Information Science, Academia Sinica

Department of Information Management, National Taiwan University

Week 14, Fall 2008

# Outline

- ▶ Regular tree grammar as an abstract model for studying XML schema languages
  - ▶ Local tree grammar, single-type tree grammar, and regular tree grammar
  - ▶ Closure property
  - ▶ Relationship to XML DTD, XML Schema, and Relax NG
  - ▶ Validation of tree structures against local/single-type/regular tree grammars
- ▶ Regular expressions and (non-)deterministic content models
  - ▶ Regular expressions and (non-)deterministic finite state automata
  - ▶ Closure property
  - ▶ Validation of element sequences against regular expressions
- ▶ Discussion