

Parsing XML into programming languages

JAXP, DOM, SAX, JDOM/DOM4J,
Xerces, Xalan, JAXB

Parsing XML

- Goal: read XML files into data structures in programming languages
- Possible strategies
 - Parse by hand with some reusable libraries
 - Parse into generic tree structure
 - Parse as sequence of events
 - Automagically parse to language-specific objects

Parsing by-hand

- **Advantages**
 - Complete control
 - Good if simple needs – build off of regex package
- **Disadvantages**
 - Must write the initial code yourself, even if it becomes generalized
 - Pretty tedious and error prone.
 - Gets very hard when using schema or DTD to validate