

Structure of the XML file for the Syntax Tree

you're near the point at which you'll have to come up with a **parser** that 'eats' the XML file that contains the tokens (from the lexer) and produces either a **syntax tree** or a "syntax error!" message.

For the same good reason as before, it is highly recommendable to **store the syntax tree permanently in a new XML file**, which can then later be "eaten" by the compiler's software modules for semantic analysis (such as type checking) and target code generation.

For this purpose I'm suggesting to you the following **XML format for the syntax tree**:

<SYNTREE>

<ROOT>

<UNID>number</UNID> *// comment: UNID means Unique Node ID number: a number which no other node has*

<SYMB>startsymbol</SYMB> *// comment: Root always contains the grammar's start symbol*

<CHILDREN>

<ID>number</ID> *// this number is the UNID of a child*

... // etc... More children ...

<ID>number</ID>

</CHILDREN>

</ROOT>

<INNERNODES>

<IN> *// comment: IN means Inner Node, between the Root and the Leafs*

<PARENT>number</PARENT> *// this number is the UNID of the parent*

<UNID>number</UNID> *// this number is the node's OWN UNID*

<SYMB>nonterminal</SYMB> *// comment: Inner nodes always contain some Nonterminal symbol of the grammar*

<CHILDREN>

<ID>number</ID> *// this number is the UNID of a child*

... // etc... More children ...

<ID>number</ID>

</CHILDREN>

</IN>

... // etc... More inner nodes ...

<IN>

<PARENT>number</PARENT>

<UNID>number</UNID>

<SYMB>nonterminal</SYMB>

<CHILDREN>

<ID>number</ID>

... // etc... More children ...

<ID>number</ID>

</CHILDREN>

</IN>

</INNERNODES>

<LEAFNODES> *// Here we have reached the level of the Token nodes that come from the Lexer*

<LEAF>

<PARENT>number</PARENT> *// this number is the UNID of the parent in the Tree*

<UNID>number</UNID> *// this number is the Leaf-node's OWN UNID in the Tree*

<TERMINAL>

... // in here you copy&paste the XML Token from the lexer as Terminal Symbol !

</TERMINAL>

</LEAF>

... // etc... More Leaf Nodes ...

<LEAF>

<PARENT>number</PARENT>

<UNID>number</UNID>

<TERMINAL>

... // in here you copy&paste the XML Token from the lexer as Terminal Symbol !

</TERMINAL>

</LEAF>

</LEAFNODES>

</SYNTREE>

The unique node ID numbers can later be used as 'Foreign Keys' in the semantic Symbol Table (see Chapter #3)