Definition of the XML-- language

Extended BNF grammar:

Document	Element
Element	StartTag (Element | Data)* EndTag | EmptyTag
StartTag	**<** Name Attribute* **>**
EndTag	**</** Name **>**
EmptyTag	**<** Name Attribute* **/>**
Attribute	Name **=** String

Lexical summary:

Name	Initial Other*
Initial	Letter | **_** | **:**
Other	Initial | Digit | **-** | **.**
String	**"** (Char | **'**)* **"** | **'** (Char | **"**)* **'**
Data	Char$^+$
Char	Ordinary | Special | Reference
Ordinary	not (**<** | **>** | **"** | **'** | **&**)
Special	**<** | **>** | **"** | **'** | **&**
Reference	**&#** (Digit)$^+$**;** | **&#x** (Digit | **a**...**f** | **A**...**F**)$^+$**;**
Letter	**a**...**z** | **A**...**Z**
Digit	**0**...**9**
Comment	**<!--** ((not **-**) | (**-** (not **-**)))* **-->**
Whitespace	blank | tab (\t) | newline (\n) | return (\r)

Additional constraints:

- The tokens are Name, String, Data, **<, >, </, />**, and **=**.
- Names and **=** may only appear inside tags, so elsewhere treat them as Data.
- Names are case-sensitive.
- Each StartTag and its corresponding EndTag must have matching names.
- Within each StartTag or EmptyTag, all attribute names must be distinct.
- The start symbol of the grammar is Document.
- Comments and Whitespace characters are ignored (except within a String).