What is XML well-formedness and how could you measure it?

Author: Mihai Rotaru

Date: 1 Nov 2011

If an XML document is well-formed, it conforms to the syntax rules, as defined in the XML standard. Unlike HTML, the XML standard is very strict, and does not tolerate errors such as the omission of a closing tag, or having different cases in the opening and closing tags; for example, the following is not a well-formed XML document:

```
1 <?xml version="1.0"?>
2 <person>
3 <name>John Smith</Name>
4 </person>
```

Although the "name" tag was closed, the closing tag tag should have the same casing as the opening tag, therefore this document is not a well-formed XML file. The line should be: <name>John Smith</name> to form a well-formed XML document.

Well-formedness also implies correct nesting; for example, the following code is not well-formed because the <surname> tag was opened before the <name> tag was closed, therefore the two tags overlap - which is not allowed in XML.

Wrong:

```
<name>John<surname>Smith</name></surname>
```

Correct:

```
<name>John</name><surname>Smith</surname>
```

Another way in which documents can fail the well-formedness test is by not having a single root tag; in other words, all of the tags inside the document must reside inside a unique pair of opening and closing tags. But it doesn't have to be the xml tag; it could be any tag (with a valid name, of course).

Well-formedness can be checked with special tools - for example, the w3scools validator (http://www.w3schools.com/xml/xml_validator.asp); but due to the strictness of the XML standard, the parser should abort at the first encountered error. Therefore, the number of errors is rarely important; what matters is whether the whole document is valid.