
Subject: Bug in the SAX parser?

Posted by [mattf](#) on Tue, 09 Jan 2007 14:34:58 GMT

[View Forum Message](#) <> [Reply to Message](#)

Just ran into a, uh, interesting problem with the SAX parser in IDL. At one point in processing an XML file, I have to remove the ignorable white space-- so I wrote a 'remove_white' subclass of the SAX parser to do that (a good exercise, btw, for anyone who's learning about this stuff). However, it turns out that the file I want to process has two-byte characters-- the 'remove_white' parser isn't bothered by that, but it appears to have a problem with recognizing a Windows newline (i.e., linefeed + carriage return) in two-byte character form. Specifically, it spots the carriage return, but not the linefeed-- so, my output file still has all the linefeeds. This is easy enough to deal with, in the subclassed 'Chars' method I test for

```
chars[0] ne string(10B)
```

before sending the string to the output file. But I think it's a bug.
