

# Removing BOM from UTF-8

---

*Source:* <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/questions/2006-02/msg01770.html>

---

- *From:* Gerard Seibert <[gerard@xxxxxxxxxxxxxx](mailto:gerard@xxxxxxxxxxxxxx)>
  - *Date:* Sat, 18 Feb 2006 11:28:18 -0500
- 

I have a large number of text files created in MS Word and saved in UTF-8 format. Unfortunately, MS Word adds the BOM to each file. I need to remove the BOM.

Information regarding BOM and UTF-8 can be found here:

<http://www.cl.cam.ac.uk/~mgk25/unicode.html>

<http://www.w3.org/International/questions/qa-utf8-bom>

A brief excerpt:

It has also been suggested to use the UTF-8 encoded BOM (0xEF 0xBB 0xBF) as a signature to mark the beginning of a UTF-8 file. This practice should definitely not be used on POSIX systems for several reasons:

- \* On POSIX systems, the locale and not magic file type codes define the encoding of plain text files. Mixing the two concepts would add a lot of complexity and break existing functionality.
- \* Adding a UTF-8 signature at the start of a file would interfere with many established conventions such as the kernel looking for #! at the beginning of a plaintext executable to locate the appropriate interpreter.
- \* Handling BOMs properly would add undesirable complexity even to simple programs like cat or grep that mix contents of several files into one.

It has been suggested that a script could be written to eliminate the BOM from a file(s). My script writing skills suck. I have been unable to locate one using Google, so I was hoping that someone might know where I could either locate such a program, or perhaps give me an idea on how to script one.

Thanks!

—

Gerard Seibert  
[gerard@xxxxxxxxxxxxxx](mailto:gerard@xxxxxxxxxxxxxx)

## Removing BOM from UTF-8

I'm interested in the fact that the less secure a man is, the more likely he is to have extreme prejudice.

Clint Eastwood

---

freebsd-questions@xxxxxxxxxxx mailing list

<http://lists.freebsd.org/mailman/listinfo/freebsd-questions>

To unsubscribe, send any mail to "freebsd-questions-unsubscribe@xxxxxxxxxxx"