# BOM

## 「带 BOM 的 UTF-8」和「无 BOM 的 UTF-8」有什么区别？网页代码一般使用哪个？

[3 条评论](http://www.zhihu.com/question/20167122) [分享](http://www.zhihu.com/question/20167122)

按票数排序[按时间排序](http://www.zhihu.com/question/20167122?sort=created)

**16 个回答**

赞同137反对，不会显示你的姓名

**[http://pic4.zhimg.com/9b4b09990_s.jpg](http://www.zhihu.com/people/lianghai)**[**梁海**](http://www.zhihu.com/people/lianghai)**，U+6211, U+7231, U+5B83**

[莫摸末](http://www.zhihu.com/people/momomo)、[王铭烨 Arthur2e5](http://www.zhihu.com/people/Arthur200000)、[刘泓伟](http://www.zhihu.com/people/liu-hong-wei-41) [等人赞同](javascript:;)

UTF-8 不需要 BOM，尽管 Unicode 标准允许在 UTF-8 中使用 BOM。  
所以不含 BOM 的 UTF-8 才是标准形式，在 UTF-8 文件中放置 BOM 主要是微软的习惯（顺便提一下：把带有 BOM 的小端序 UTF-16 称作「Unicode」而又不详细说明，这也是微软的习惯）。  
BOM（byte order mark）是为 UTF-16 和 UTF-32 准备的，用于标记字节序（byte order）。微软在 UTF-8 中使用 BOM 是因为这样可以把 UTF-8 和 ASCII 等编码明确区分开，但这样的文件在 Windows 之外的操作系统里会带来问题。  
  
「UTF-8」和「带 BOM 的 UTF-8」的区别就是有没有 BOM。即文件开头有没有 U+FEFF。  
  
UTF-8 的网页代码不应使用 BOM，否则常常会出错。这是一个小例子： [为什么这个网页代码 <head> 内的信息会被浏览器理解为在 <body> 内？](http://www.zhihu.com/question/20138814)

## 为什么这个网页代码 <head> 内的信息会被浏览器理解为在 <body> 内？

另外，页面总是莫名出现多余的引号……地址：[http://letwindgo.com](http://letwindgo.com/) 我是新手，谢谢大家

[1 条评论](http://www.zhihu.com/question/20138814) [分享](http://www.zhihu.com/question/20138814)

按票数排序[按时间排序](http://www.zhihu.com/question/20138814?sort=created)

**2 个回答**

赞同20反对，不会显示你的姓名

**[http://pic2.zhimg.com/9b4b09990_s.jpg](http://www.zhihu.com/people/lianghai)**[**梁海**](http://www.zhihu.com/people/lianghai)**，文字设计（typography）与语音/音系爱好者**

[任继敏](http://www.zhihu.com/people/pachydermx)、知乎用户、知乎用户 [等人赞同](javascript:;)

<head> 嵌套错误是因为整个 HTML 文档的最开头多了一个看不见的 U+FEFF（就在「<!DOCTYPE」前面），它污染了 HTML。  
  
U+FEFF 就是我们常说的「BOM（byte order mark）」，也是 @董鹏 说的「signature」，UTF-8 编码的文件不需要它。Windows 该死的记事本有个臭名昭著的破毛病就是在 UTF-8 文件开头加 BOM，所以**不要用**记事本来编辑文件。  
可以用这个在线工具检查你的网页是否被多余的 BOM 污染了： <http://people.w3.org/rishida/utils/bomtester/>