

转化为 <file2>。ed 是一个非常简单的编辑器,支持一些命令。在图 15-10 的例子中,5c 说明要删除基线文档的第 5 行,而 chisels.<cr> 说明要添加 chisels.,就这么简单。对于更大的改动,会产生更复杂的指令。Unix 系统的 diff-e 算法是对文件进行逐行比较的,这对于文本文件没问题,但并不适合二进制文件。vcdiff 算法更强大,对于非文本文件也适用,并且产生的差异比 diff-e 要小。

差异编码的规范中详细定义了 A-IM 和 IM 首部的格式。在这里,我们只要知道这些首部中可以说明多个实例操控(并可以带有相关的质量值)就够了。在返回给客户端之前,文档可以经过多种实例操控,这样可以获得最大程度的压缩。例如,用 vcdiff 算法产生的差异随后可以再用 gzip 算法压缩。于是服务器的响应中就含有 IM:vcdiff, gzip 首部。客户端应当先对内容进行 gunzip,再把得到的差异应用到自己的基线页面上,这样才能生成最终的文档。

差异编码可以减少传输次数,但实现起来可能比较麻烦。设想一下页面改动频繁,而且有很多不同的人都在访问的情形。支持差异编码的服务器必须保存页面随时间变化的所有不同版本,这样才能指出最新版本与所请求的客户端持有的任意版本之间的差异。(如果文档变化频繁,而且有很多客户端都在请求文档,那它们就会获得文档的不同实例。随后当它们再向服务器发起请求时,它们将请求它们所持有的版本与最新版本之间的差异。为了能够只向它们发送变化的部分,服务器必须保存所有客户端曾经持有过的版本。)要降低提交文档时的延迟时间,服务器必须增加磁盘空间来保存文档的各种旧的实例。实现差异编码所需的额外磁盘空间可能很快会将减少传输量获得的好处抵消掉。

368

15.11 更多信息

关于实体和编码方面的更多信息,请参考以下资源。

- <http://www.ietf.org/rfc/rfc2616.txt>
RFC 2616,也就是 HTTP/1.1 版的规范,是实体主体管理和编码方面的主要参考。
- <http://www.ietf.org/rfc/rfc3229.txt>
RFC 3229,“Delta Encoding in HTTP”(“HTTP 中的差异编码”),说明了如何通过扩展 HTTP/1.1 来支持差异编码。
- *Introduction to Data Compression*¹¹ (《数据压缩导论》)
这本书的作者是 Khalid Sayood,出版商为 Morgan Kaufmann Publishers。该书介绍了几种 HTTP 内容编码支持的压缩算法。

注 11: 本书影印版由人民邮电出版社出版。(编者注)