命中比例），[4]有时也被称为文档命中率（document hit rate）。命中率在0到1之间，但通常是用百分数来描述的，0%表示每次请求都未命中（要通过网络来获取文档），100%表示每次请求都命中了（在缓存中有一份副本）。[5]

缓存的管理者希望缓存命中率接近100%。而实际得到的命中率则与缓存的大小、缓存用户兴趣点的相似性、缓存数据的变化或个性化频率，以及如何配置缓存有关。命中率很难预测，但对现在中等规模的Web缓存来说，40%的命中率是很合理的。缓存的好处是，即使是中等规模的缓存，其所包含的常见文档也足以显著地提高性能、减少流量了。缓存会努力确保将有用的内容保存在缓存中。

## 7.5.3 字节命中率

由于文档并不全是同一尺寸的，所以文档命中率并不能说明一切。有些大型对象被访问的次数可能较少，但由于尺寸的原因，对整个数据流量的贡献却更大。因此，有些人更愿意使用字节命中率（byte hit rate）作为度量值（尤其那些按流量字节付费的人！）。

字节命中率表示的是缓存提供的字节在传输的所有字节中所占的比例。通过这种度量方式，可以得知节省流量的程度。100%的字节命中率说明每个字节都来自缓存，没有流量流到因特网上去。

文档命中率和字节命中率对缓存性能的评估都是很有用的。文档命中率说明阻止了多少通往外部网络的Web事务。事务有一个通常都很大的固定时间成分（比如，建立一条到服务器的TCP连接），提高文档命中率对降低整体延迟（时延）很有好处。
167 字节命中率说明阻止了多少字节传向因特网。提高字节命中率对节省带宽很有利。

## 7.5.4 区分命中和未命中的情况

不幸的是，HTTP没有为用户提供一种手段来区分响应是缓存命中的，还是访问原始服务器得到的。在这两种情况下，响应码都是200 OK，说明响应有主体部分。有些商业代理缓存会在`Via`首部附加一些额外信息，以描述缓存中发生的情况。

客户端有一种方法可以判断响应是否来自缓存，就是使用`Date`首部。将响应中`Date`首部的值与当前时间进行比较，如果响应中的日期值比较早，客户端通常就可

注4：术语“命中比例”可能比“命中率”要好，因为“命中率”会让人错误地想到时间因素。但是“命中率”这个词很常用，所以这里我们也使用它。

注5：有时，人们会在命中率中包括再验证命中，但有时候命中率和再验证命中率是分别测量的。在检测命中率的时候，要确定自己知道什么才是“命中”。