曼哈顿相似度

1. 读取数据：矩阵格式 user\_movie\_rating，（train Set、Test Set）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Movie1 | Movie2 | Movie3 | Movie |
| User1 |  |  |  |  |
| User2 |  |  |  |  |
| User3 |  |  |  |  |
|  |  |  |  |  |

（2）

对于一个网站的结构来说，URL的设计也非常的重要，在URL设计时我们应该事先做好规划，应该从用户体验和搜索引擎蜘蛛体验出发，在总体上边应该做到URL清晰、友好、方便记忆，然后还要适当的考虑对于搜索引擎排名的影响，具体来说，URL设计时我们要从下边几点来着手：

1、URL越短越好

这样做主要是为了用户体验，其实对于搜索引擎来说，不超过1000个字母都没有问题，可是如果不考虑用户体验，确实真的用了非常长的URL的话，用户看起来就有些麻烦了，甚至会马上离开你的网站，曾经有人做过这方面的试验，在搜索引擎的搜索结果中，一个较短的URL出现在一个较长的URL下边时，较短的URL的点击率是较长的URL的2.5倍。而且较短的URL也比较有利于复制与传播。人们都比较喜欢简单的东西，特别是现代，人们不喜欢太复杂的东西，能简单就简单。所以网站的URL要尽量的短。

2、URL中避免太多参数

在进行URL设计时，我们尽量使用静态的URL，虽然现代的主流搜索引擎都能够很好的去索引带有参数的动态网页了，可是，使用静态的URL还是有一定的优势的，如果由于某些原因，不能使用静态的URL，而不得不使用动态的URL时，设计的则是URL中的参数应该尽量减少，不要超过三个，一般的情况下URL中的参数2-3个就可以了。参数太多了不但用户看起来困难，对搜索引擎的索引可能也会造成一定的问题。所以，能使用静态就使用静态的URL，能短就尽量短。

3、目录层次尽量少

这里所指的目录层次是指物理目录结构，而不是指逻辑结构，我们在进行URL的设计时，网站的结构要尽量的去减少目录层次，层次不能太深了，一般建议不要超过三层，特别对于一些新站来说，权重低，搜索引擎蜘蛛爬行得很浅，深一点的页面，蜘蛛都很可能不会去爬行的，所以要尽量的做到使目录层次减少，URL缩短。根据观察，百度尤其比较喜欢目录层次比较少的页面。

4、文件名及目录名要具描述性

文件名及目录名要具有可描述性，不但让用户一眼就能看出来这个页面是关于什么的，对用户体验比较友好，而且搜索引擎也比较喜欢这样的URL。例如一个关于新闻的目录，我们可以把它命名为news，用户看到这个目录名称，大概就知道这个目录是关于什么内容的了。

5、URL中包含关键词

关键词出现在URL中，可以提高一定的相关性，尤其对于优化英文网站，这点非常的重要，而且关键词出现在URL的位置越靠前越好，也就是说出现在域名中最好，其次是出现在目录名中，最后才是出现在文件名中。注意一点的是不过不要出现关键词堆积的现象。例如许多朋友的SEO博客，在网站的域名中就包含了SEO，这对于关键词的排名来说，具有一定的权重优势。

6、URL中的字母全部用小写

这里有以下原因：

(1)、全部用小写，用户比较容易输入，不用因为大小写混合而出现错误，这是人们的输入习惯。

(2)、有些服务器是区分大小写的，例如Lunix服务器，这样在站长做链接或者是用户输入时，会因为大小写的问题而出现404错误，而且robots也是区分大小写的，如果大小写搞错了，可能会造成不能收录的严重问题。所以建议所有的URL都使用小写。

7、连词符的使用

目录或者文件名中如果有两个单词组成时，一般建议中间使用中划线(-)隔开，切记不要使用下划线或者其他字符，在搜索引擎中，它是把中划线当作一个空格来处理的，而下划线则是被忽略的，例如seo-lunwen会被读成seo与lunwen。这是比较友好的写法。

8、使用目录形式还是文件形式

大部分的CMS系统都可以把页面的URL设置为目录或者文件形式，有些朋友认为目录形式比文件形式权重要高一点，无论使用哪种形式的URL，权重应该是差不了多少的，基本可以忽略二者之间的权重差异，使用目录形式有一个优点，以后如果更换其他的编程语言，URL也可以不用变化，也不用经过特殊的处理，而文件形式的URL可能文件的扩展名会变化，需要做一定的URL重写处理。

上边是一个网站的URL在设计的时候，应该注意的八个问题。在设计URL时，只有严格的按照这几点来操作就可以了。网站的URL设计得好，不但利于蜘蛛的爬行，而且用户看着也简单、易懂。