

HOMEWORK-Part2-Leo

Google搜索引擎是如何对搜索结果进行排序的？（请用你自己的语言描述PageRank算法。）

PageRank算法顾名思义就是网页排名算法，它主要是通过网络之间的超链接关系来确定一个页面的等级，等级越高，说明页面越受欢迎，Google工具条上的PageRank等级划分是从0到10，10级为满分，比如一个网站的PR值为1表明这个网站不太具有流行度，而PR值为7到10则表明这个网站非常受欢迎。

我们可以把PageRank算法理解成投票打分，我们把从A页面到B页面的链接理解成A页面给B页面投票，Google根据投票来源和投票目标的等级来决定新的等级，简单讲，一个高等级的页面可以使得其他低等级的页面等级提升。

PageRank就是让链接来投票，一个页面的得票数由所有链向它的页面的重要性来决定，到一个页面的超链接相当于对该页投一票。一个页面的PageRank是由所有它的链入页面的重要性经过递归算法来得到的。一个有较多链入的页面会有较高的等级，相反如果一个页面没有任何链入页面，那么它没有等级。

这里举个例子，加入一个由4个页面组成的小团体：A，B，C，D。如果所有页面都链向A，那么A的PageRank等级值时B，C，D的PageRank总和：

$$PR(A) = PR(B) + PR(C) + PR(D)$$

继续假设B也有链接到C，并且D也有链接到包括A的3个页面。一个页面不能投票两次，所以B给每个页面半票。同样的逻辑，D投出的票只有1/3算到了A的PageRank上。

$$PR(A) = PR(B)/2 + PR(C)/1 + PR(D)/3$$

换句话说，根据链出总数平分一个页面的PR值。公式如下：

$$PR(A) = PR(B)/L(B) + PR(C)/L(C) + PR(D)/L(D)$$

最后，所有这些被换算为一个百分比再乘上一个系数。由于“没有向外链接的页面”传递出去的PageRank会是0，所以，Google通过数学系统给了每个页面一个最小值：

$$PR(A) = (PR(B)/L(B) + PR(C)/L(C) + PR(D)/L(D) + \dots) * d + (1-d)/N$$

以上即是PageRank算法的简要描述。

