1、PageRank

1.1.简介

PageRank，又称网页排名、谷歌左侧排名，是一种由搜索引擎根据网页之间相互的超链接计算的技术，而作为网页排名的要素之一，以Google公司创办人拉里·佩奇（Larry Page）之姓来命名。Google用它来体现网页的相关性和重要性，在搜索引擎优化操作中是经常被用来评估网页优化的成效因素之一。

假设一个由4个网页组成的群体：A，B，C和D。如果所有页面都只链接至A，那么A的PR（PageRank）值将是B，C及D的Pagerank总和。

IMG_256

重新假设B链接到A和C，C只链接到A，并且D链接到全部其他的3个页面。一个页面总共只有一票。所以B给A和C每个页面半票。以同样的逻辑，D投出的票只有三分之一算到了A的PageRank上。

IMG_257

1.2.公式

对于一个页面A，那么它的PR值为：

IMG_258

PR(A) 是页面A的PR值

PR(Ti)是页面Ti的PR值，在这里，页面Ti是指向A的所有页面中的某个页面

C(Ti)是页面Ti的出度，也就是Ti指向其他页面的边的个数

d 为阻尼系数，其意义是，在任意时刻，用户到达某页面后并继续向后浏览的概率，

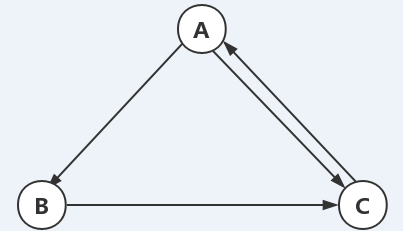
该数值是根据上网者使用浏览器书签的平均频率估算而得，通常d=0.85

还有一个版本的公式：

IMG_259

N为页面的总数

1.3.具体实例



三个页面A、B、C

为了便于计算，我们假设每个页面的PR初始值为1，d为0.5。

页面A的PR值计算如下：

IMG_261

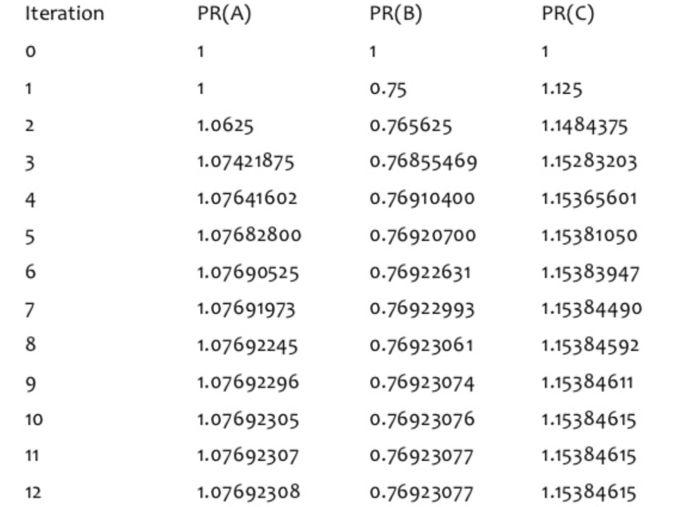
页面B的PR值计算如下：

IMG_262

页面C的PR值计算如下：

IMG_263

下面是迭代计算12轮之后，各个页面的PR值：



那么什么时候，迭代结束哪？一般要设置收敛条件：比如上次迭代结果与本次迭代结果小于某个误差，我们结束程序运行；比如还可以设置最大循环次数。