PageRank学习日志

7.2：

PageRank的核心思想很简单，主要基于以下两点：

1. 如果一个网页被很多其他网页链接到的话说明这个网页比较重要，也就是PageRank值会相对较高
2. 如果一个PageRank值很高的网页链接到一个其他的网页，那么被链接到的网页的PageRank值会相应地因此而提高

PageRank算法初始会预先给每个网页一个PR值，PR值物理意义上可以简单理解为一个网页被访问到的概率，一般为1/N，其中N为网页总数

预先给定PR值后，通过转换算法不断迭代，直至达到PR值趋于稳定

在一般情况下，一个网页的PR值计算如下：

*PR*(*pi*)=*α*∑*pj*∈*MpiPR*(*pj*)*L*(*pj*)+(1−*α*)*N*

其中*Mpi*是所有对*pi*网页有出链的网页集合，*L*(*pj*)是网页*pj*的出链数目，*N*是网页总数，*α*一般取0.85。  
 根据上面的公式，我们可以计算每个网页的PR值，在不断迭代趋于平稳的时候，即为最终结果。