

why is this paper named “topic-sensitive”?

为了得到更加精确的结果，文章中的网页搜索算法使用具有代表性的相关主题对 PageRank 向量进行便宜，以更准确地捕捉关于特定主题的相关性，对于上下文中进行的搜索，文章使用查询出现的上下文主题来计算主题敏感的 PageRank 分数，这样就可以生成比单纯通用的 PageRank 向量更准确的排序。

What’s the problem this paper was trying to address?

普通的 PageRank 预先计算出一个等级向量，给所有的 document 一个先验的重要性估计，这个估计与查询无关，只是通过了网页之间的连接结构来判断重要性，忽略了查询词的相关性。由于一些网页在特定的话题下很重要，而并不一定和 PageRank 的重要性结果呈现完全的相关性。这些在传统 PageRank 认为很重要的网页可能并不能和查询话题很好地契合。

What’s the role of matrix D?

$$D = \vec{p} \times \vec{d}^T$$

矩阵 D 用于指示某个网页的出度是否为 0，如果是，那么对应项为 1。这些出度为 0 的网页将按照 \vec{p} 中指示的概率访问其他网页。公式中的概率矩阵为 $(M+D)$ ，其中 M 指示网页链接结构中得到的概率关系，当某个网站的出度为 0，那么其对应的列中全为 0；加上矩阵 D，实际上考虑到了两个方面——网页的链接结构和没有出度的网页，这样就在随机游走的模型中弥补了出度为 0 的网页的情况。

Why we need the ODP or other similar project for the computing of multiple page ranks for a given page?

ODP 数据是免费提供的，而且由于它是由数千名自愿编辑汇编的，因此不太容易受到任何一方的影响。传统的 PageRank 算法由于被设置成话题敏感的，需要对计算公式增加一些偏向，首先需要利用爬虫得到一组对应于各个话题的网页集合，那么 ODP 就是这样一个实用、可靠且免费的信息来源。