**2.3.1搜索排序因子的分析**

社交网络的主要特征是以用户为中心，用各种关系组成了一个小的社交关系网。对用户关系进行分析评估，挖掘可能影响搜索排序的因子至关重要；同时，亦要考虑社交网络中的信息的实时性，以及附加信息等等。所以对用户搜索排序主要考虑的因子有以下几点。

1)朋友权威度：

基于基本假设假：设用户权威度越高，代表其发布内容越具影响力和价值。通过一定特征描绘朋友的权威性，并加以利用往往能提升搜索排序质量。而根据社会学，一个人在该社交网络中拥有越多朋友，其威望度往往越高；而一个人的朋友威望越高，那么他的威望往往也会越高。据此我们将采用PageRank算法进行权威度计算。

2)好友亲密度：

根据社交网络的特点，用户更多的想关注并信赖与其关系较密切的好友所发布的内容，因此引入亲密度作为排序生成的因子之一。亲密度的衡量将充分考虑好友互动内容，互动计数，共同好友数量等等。

3）附加信息：

社交网络中往往存在大量附加信息，包括博文的长度，发布时间，评论数，转发数，点赞数等等，这些因子都一定程度上反映了博文的价值，可作为附加的搜索排序因子，优化搜索排序算法。

4)缓存机制：

经统计发现搜索关键字查询的集中现象非常明显，20%的关键字占到总查询次数80%。因此建立缓存机制，对20%左右关键字进行缓存，能加快搜索排序的速度。

**2.3.3基于PageRank改良算法的好友权威度分析**

PageRank算法，是由斯坦福大学的Larry Page和Sergey Brin提出来的，也是Google网络搜索引擎采用的检索排序算法之一。PageRank算法基于一个基本假设：“从许多优质网页链接过来的网页，必定还是优质网页。”因此，算法的基本思路是通过计算指向一个页面中的链接数量和质量来决定这个页面的重要性。同样，在社交网络，如果一个有一个朋友，相当于该朋友给他投了一票，如果该用户有很多的朋友，说明他在该社交网络中的具有很高的威望，就像是在微博中，粉丝量高的往往是那些重要人物或者媒体。同样，一个人如果有一个朋友威望很高，说明他的威望也很高。所以根据这个特点，我们可以利用PageRank算法进行好友权威度的计算。

式中，PR（v）表示发表帖子的用户v的权威度，d表示衰减因子，其取值介于0~1之间，通常设定为0.8，指用户v的朋友，PR（u）表示朋友u的权值，表示朋友的个数。

**2.3.4附加信息分析**

对于社交网络数据价值的判断，往往不单单只是考量数据内容本身。他还有许多其他附加信息影响数据的价值。以微博数据为例，用户所发布博文所附带的时间，博文长度；以及博文的转发数，评论数等等，都可作为数据价值的参考因素。因此有必要对附加信息进行分析，并加以利用。

a)时间差：

社交网络的数据一般集中在对当前的社会热点话题，其时间属性相比于传统网页数据具有更大的价值。现在很多突发性信息也是通过社交网络发布出去，它比很多传统媒体都要迅捷。从这点可以看出，社交网络数据的发布时间，对信息价值具有较大影响。一般认为用户搜索时的时间与信息发布的时间之差越小，其价值越大。

b)信息长度：

社交网络的信息长度相对短小，例如国外的facebook，和国内的微博，都对发布信息有严格的长度限制。其所包含的信息量有限，研究表明，越长的社交网络数据具有越大的价值，在最后的排序结果的排名应该越靠前。

c)评论数和转发数：

由于社交网络信息发布门槛低，几乎每个人都可在上面发布信息，而且信息发布的周期很短，导致了网络中信息价值密度低的特点。一般认为，一篇博文的评论数以及转发数越高，表明其价值更高，其排名应该越靠前。

发布时间

信息长度

评论数

转发数

反比

正比比

正比比

正比比

信息价值

正比比

搜索排名靠前