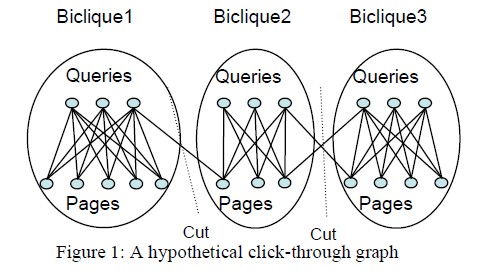
1.【query clustering using click-through graph】通过使用click-through graph来表示用户查询以及用户点击过的文本的关系，在生成连接图的过程中通过剪枝策略消除不符合条件的query节点和page节点，具体聚类过程如下：

首先，构建click-through graph，将用户提交的查询节点与点击过的page节点连接起来：



然后，由click-through graph递归的构造query-pages二元组：

1. 从queries序列中选出query qk ,以及它链接到的页面数i，并找出它链接到的所有页面P(qk)
2. 对于每一个P(qk)，找到所有链入它的queries
3. 计算出上一步中所有queries链入的pages ∩l∈p(qk)Q(pi)
4. 假设第2步中求出的queries数为m，E是所有queries链入到p(qk)的边，如果（m≧j）,构造一个大小为（i,m）的二元组，然后移除所有的query、page节点和边，如果（m<j）删除所有从第2步中找到的query节点到p(qk)的边
5. 删除所有边以后，执行剪枝策略并进行下一步迭代

最后，生成所有二元组以后，对于每一个二元组中的queries构成一个聚类

2.【query clustering using a hybrid query similarity measure】结合了content-based和result-based两种方法进行用户查询的聚类，content-based方法是指根据用户提交的查询词本身判断查询之间的相关性，即当两个查询词之间有相同的term时，说明这两个查询词之间存在相关关系，result-based是指根据用户提交查询后返回的结果url中重复的文档数计算用户查询之间的相关性。本文将两种方法线性结合在一起，在互相弥补各自缺陷的基础上可以达到聚类整体性能的最好，其中content-based使用tf-idf计算查询词的权值，使用cosin公式计算查询词之间的相关性，而result-based方法中使用两个查询词返回相同的文档数计算其相关性，最后将两者线性结合在一起计算最终的相关性。