各位老师好，我叫刘大力。我的学位论文题目是《面向搜索经验的查询推荐方法研究》。我的导师是张斌老师。

今天的答辩我想从研究背景、研究思路与研究方法三个方面介绍我的研究。

首先介绍一下复杂搜索，复杂搜索描述了用户在进行搜索过程中的3种典型场景：系统对信息的索引不够充分、搜索任务本身需要浏览及探索以及用户难以对查询进行组织或定位信息领域。

复杂搜索的特性决定了用户的复杂搜索过程常常具有：周期长、间断多以及信息负载重的特点。

传统的查询推荐方法不能很好地利用复杂搜索特性进行查询推荐，因此我们希望提出一种针对复杂搜索的查询推荐方法，以帮助用户在复杂搜索过程中更高效地达成信息需求。

在此前的研究中我们我们实验室提出了时间树理论帮助用户管理复杂搜索过程。时间树将用户在复杂搜索过程中进行的查询用圆圈表示，将用户进行的点击用方块表示。并将他们组织称一棵树的形式。

由于时间树的结构特性，用户从时间树中能够很容易地获取到以下信息：

节点产生的相对时间顺序。

查询的来源。

子任务的划分。

我们猜想，既然时间树能够帮助用户获得这些信息，很有可能是因为时间树中蕴含了用户在复杂搜索过程中产生的高质量的搜索经验。于是就有了我们这个，面向搜索经验的查询推荐方法研究。

那么本研究的研究思路是这样的：作为本研究的基础，首先我们验证时间树中蕴含高质量的搜索经验，然后我们研究如何提取搜索经验，再搞定如何利用提取的搜索经验进行查询推荐。最后我们将理论应用于实践，设计实现面向搜索经验的查询推荐系统。

我们定义了搜索经验模型，将用户在复杂搜索过程中的搜索经验分为时间经验、因果经验以及主题经验。

时间经验为用户在搜索过程中进行查询和点击的时间先后关系。

因果经验为用户所进行查询的查询原因。

主题经验为用户对复杂搜索任务所进行的子任务划分。

在定义了搜索经验模型的基础上，我们设计实验，从用户学习领域经典的主观评估、专家评估与客观评估三个角度，验证了时间树中蕴含有高质量的因果经验以及主题经验。这是实验结果。

那么下面我们就可以分别针对因果经验与主题经验研究提取的方法。

于是就有了基于查询-点击-查询序列识别的因果经验提取算法，和基于子任务划分的主题经验提取算法。

因果经验提取算法就是将时间树中的查询-点击-查询序列提取出来得到因果经验的集合。这是算法伪代码。

主题经验提取算法首先将时间树上的节点网格化，然后为每个节点设置以网格边长2倍为边长的范围，如果两节点的范围有重合则将他们聚为一类，如果不重合则不聚为一类，这样可以将时间树上的节点划分为不同子任务，以此提取主题经验。这是算法为代码。

我们设计实验验证了主题经验提取算法的有效性，这是实验结果。

然后我们就可以想办法利用搜索经验进行查询推荐了，面向搜索经验的查询推荐方法分为面向因果经验的查询推荐方法、子任务内部的查询推荐方法以及跨子任务的查询推荐方法三种。

在具体介绍这三种方法之前，我想我们来看一下查询推荐方法和搜索经验提取方法的对应关系。面向因果经验的查询推荐方法利用提取的因果经验进行查询推荐，子任务内部的查询推荐方法和跨子任务的查询推荐方法都面向主题经验。只不过面向主题经验查询推荐之前需要先进行主题经验的合并处理，因此这里要加一层主题经验合并算法。

面向因果经验的查询推荐方法是将用户提交的查询，与因果经验集合中的查询-点击-查询序列进行匹配和筛选，得到匹配度最高的查询-点击-查询序列进行推荐。

面向主题经验的查询推荐首先要进行主题经验合并，将提取出的每一个用户的子任务子树集合分别合并为含有用户所提交查询的子任务子树和不含有用户所提交查询的子任务子树集合。

含有用户所提交查询的子任务有向图用于子任务内部的查询推荐，不含有用户所提交查询的子任务有向图集合用于跨子任务的查询推荐。

子任务内部的查询推荐方法首先计算很有用户所提交查询的子任务有向图中所有节点的rank值，取出rank值最高的节点作为推荐节点，然后计算从用户提交的查询节点到推荐节点的最优路径。

跨子任务的查询推荐方法针对不含有用户所提交查询的子任务有向图集合中的每一张子任务有向图，计算出rank值最高的节点作为推荐节点，再计算求得从推荐节点出发，rank值最高的几条路径作为推荐路径。

同样我们也设计实验验证了面向搜索经验的查询推荐方法的有效性。这是实验结果。

在上述理论的基础上我们设计并实现了面向搜索经验的查询推荐系统。这是系统架构和系统实例。

我的讲解内容就是这些，谢谢各位老师。