# 第三章

3.1 时间树与用户的搜索经验

在此前对复杂搜索的研究中，作者所在实验室提出了时间树理论来对复杂搜索任务进行管理，并对复杂搜索过程进行可视化。在时间树理论中，每一个搜索任务都可以被表示成一棵时间树。在时间树中，查询和点击被表示为节点，并且被组织成一种带有时间关系的源追踪树型结构（relative chronological source-tracking tree ，后文中简称RCST结构）。

时间树的数学模型。

这种结构是时间树的最典型特征。基于RCST结构，用户可以快速地分辨查询与点击之间的关系。例如，用户可以根据一些节点的位置关系来判别这些节点间的相对时间顺序，也可以从一个查询的源节点得到查询动机的启发，用户还可以轻松地根据时间树的RCST结构对非常复杂的搜索任务进行子任务的划分。

时间树能够记录用户的搜索过程，我们有理由提出假设，时间树中蕴含了用户的搜索经验。本文希望提出一种基于搜索经验的查询推荐方法，并依据此方法设计并实现查询推荐系统以形成一套完整的基于搜索经验的探索式搜索查询推荐解决方案。

为了更好地验证时间树中蕴含用户搜索经验这一假设，我们首先需要对用户的搜索经验进行建模。因此我们提出了搜索经验模型。我们将搜索经验模型定义为一个三元组E<TE, CE, THE>，其中TE，CE和THE分别表示用户在进行探索式搜索过程中的时间经验，因果经验以及主题经验。其中时间经验为用户在搜索中进行查询点击行为的时间先后关系，例如用户首先查询了某个词，进行了某些点击，又查询了某些词。因果经验为用户所进行查询的查询来源，例如用户从某篇文章中看到了某个关键词获得了查询灵感，或者用户是进行了一次查询重构。用户在进行一次复杂搜索任务的过程中，经常会将复杂的搜索任务划分为几个子话题、子任务。这些子任务提供了一种复杂搜索任务的清晰结构展现，反映了用户对其所进行的复杂搜索任务的规划。我们就将用户对子任务的划分定义为主题经验。

搜索经验模型将用户的搜索经验划分为时间经验，因果经验以及主题经验三种经验，我们就可以通过分别验证时间树中对这三种搜索经验的蕴含性来验证时间树中对用户搜索经验的蕴含性。但是用户的这三种搜索经验并不能直观地观测到，为了解决这一问题，我们使用用户对搜索过程的回想来从侧面验证这三种搜索经验。对过去事件的回想对于解决问题的重要性并不亚于其对于社会交往的重要性。例如，人们可能会需要向他的同事介绍他解决一个复杂问题时的搜索过程，或者可能会需要在解决新任务的时候，应用一些以往的搜索经验。想要表述过去的搜索任务，人们需要有条理地对记忆进行组织。Habermas等人提出了人们对过去事件回想的评价标准，好的回想应当具有三个性质：时间一致性，因果一致性以及主题一致性。我们可以通过分别验证使用时间树的用户在对搜索过程的回想中的这三种一致性来评价时间树对时间经验，因果经验以及主题经验的蕴含性。

因此本章将从用户对复杂任务搜索过程的回想的角度设计实验，通过对比用户在使用时间树与使用普通线形日志搜索后，依据两种不同日志形式对搜索过程进行回想时在时间一致性，因果一致性以及主题一致性上的表现来评价时间树与普通线性日志对时间经验，因果经验以及主题经验的蕴含性。

## 3.1 搜索经验模型

### 3.1.1 用户学习

为了建立用户搜索经验模型，我们首先要分析用户在搜索过程中都会产生哪些经验。我们对用户进行复杂搜索的过程进行分析。最基本也最容易观察到，用户的查询和点击会按照时间先后进行，用户不会同时进行两个查询或点击行为，因此用户的查询和点击一定具有一定的时间先后顺序。其次，用户进行某个查询，一定是具有一定的原因，例如来自点击结果，或者来自用户当前所进行任务的任务需求，或者来自突然迸发的灵感。最后，用户在进行复杂搜索任务的过程中，由于任务的复杂性，用户常常需要将任务划分为各个小的子任务，并在逐一完成各个子任务的过程中完成复杂搜索任务。子任务的划分对复杂搜索任务的完成有很重要的意义，它反映了了用户对搜索任务理解的完整度，更好地对子任务进行划分，有助于用户更清晰地完成复杂搜索任务。

但用户在进行搜索的过程中，每一个子任务并不是连贯完成的，用户常常会因为各种干扰因素而发生子任务的切换，例如用户在进行子任务A的过程中点击了一篇文章，文章中提到了一个关键词C，而这个关键词是属于子任务B中的内容，用户如果搜索了关键词C，便产生了一次由子任务A到子任务B的切换，而此时，子任务A尚没有被完成，在未来的某个时间点，用户很可能会重新对子任务A进行搜索。因此，子任务往往散落在用户搜索过程中的各个时间点上，因此，将离散杂乱的搜索过程划分为各个子任务的过程，也应该是用户产生的搜索经验中很重要的一点。

据此，我们观察到了在用户搜索过程中产生的各种搜索经验的表现，接下来我们将依据这些观察，提出搜索经验模型。

### 3.1.2 搜索经验模型

我们将用户在进行复杂搜索过程中的搜索经验定义为一个三元组E<TE, CE, THE>，其中TE为时间经验，CE为因果经验，THE为主题经验。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，即在所有用户进行的查询和点击中的任意两个所具有的时间先后关系。因果经验为用户在搜索过程中进行某个查询的原因。主题经验为用户对复杂搜索任务所进行的子任务划分。

时间经验，因果经验，主题经验共同组成了用户在探索式搜索过程中的搜索经验模型，但是这三种用户的搜索经验都无法直观地观测到或是直接进行衡量。为了解决这一问题，我们考虑使用某种侧面的评价标准，来衡量用户的搜索经验。我们提出一种合理的推断：如果用户在搜索过程中产生了搜索经验E<TE, CE, THE>，并且某种日志形式能够记录这种搜索经验，那么用户在依据这种日志形式回忆自己的搜索过程时，一定能够分别满足时间一致性，因果一致性，主题一致性。这三种一致性的概念由Habermas等人在2000年提出，他们指出，好的回想应当具有这三种性质。其中时间一致性指用户的能够准确回忆起自己在搜索过程中查询与点击的先后顺序。因果一致性指用户能够准确回忆起搜索中的某个查询的查询动机。主题一致性指用户能够合理地对任务进行子任务划分，以及对搜索过程中的某一个阶段，能够判断其中的查询与点击分别属于哪些子任务。这三种一致性与搜索经验模型中的TE, CE, THE相对应。如果某种搜索日志形式能够很好地帮助用户记录搜素经验E<TE, CE, THE>，那么用户在依据这种搜索日志进行搜索过程回想时一定能够获得更好的时间一致性，因果一致性以及主题一致性，反之，如果一种搜索日志形式不能很好地记录用户的搜索经验，那么用户在依据这种搜索日志形式进行搜索过程的回想时一定不能获得很好的时间一致性，因果一致性和主题一致性。

## 3.2 实验