分类号 密 级 U D C

学 位 论 文

|  |  |  |  |
| --- | --- | --- | --- |
| 作者姓名： | 刘大力 | | |
| 指导教师： | 张斌 教授 | | |
|  | 东北大学计算机科学与工程学院 | | |
| 申请学位级别： | 硕士 | 学科类别： | 工学 |
| 学科专业名称： | 计算机应用技术 | | |
| 论文提交日期： | 2017年10月 | 论文答辩日期： | 2017年12月 |
| 学位授予日期： | 2017年10月 | 答辩委员会主席： |  |
| 评阅人： |  | | |

东 北 大 学

2017年10月

**A Thesis in Computer Software and Theory**

By Liu Dali

Supervisor: Professor Zhang Bin

**Northeastern University**

October 2017

独创性声明

本人声明，所呈交的学位论文是在导师的指导下完成的。论文中取得的研究成果除加以标注和致谢的地方外，不包含其他人己经发表或撰写过的研究成果，也不包括本人为获得其他学位而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均己在论文中作了明确的说明并表示谢意。

学位论文作者签名：

日 期：

学位论文版权使用授权书

本学位论文作者和指导教师完全了解东北大学有关保留、使用学位论文的规定：即学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人同意东北大学可以将学位论文的全部或部分内容编入有关数据库进行检索、交流。

作者和导师同意网上交流的时间为作者获得学位后：

半年 □ 一年□ 一年半□ 两年□

学位论文作者签名： 导师签名：

签字日期： 签字日期：

# 摘 要

关键词：

# Abstract

**Key words:**

目 录

[独创性声明 I](#_Toc488500596)

[摘 要 II](#_Toc488500597)

[Abstract III](#_Toc488500598)

[第1章 引 言 - 1 -](#_Toc488500599)

[第2章 相关研究 - 2 -](#_Toc488500600)

[第3章 - 3 -](#_Toc488500601)

[3.1 时间树与用户的搜索经验 - 3 -](#_Toc488500602)

[3.2 用户搜索经验模型及用户搜索经验一致性模型 - 6 -](#_Toc488500603)

[3.3 实验设计 - 8 -](#_Toc488500604)

[3.3.1 复杂搜索任务设计 - 8 -](#_Toc488500605)

[3.3.2 实验过程设计 - 9 -](#_Toc488500606)

[3.4 实验结果分析 - 12 -](#_Toc488500607)

[3.4.1 主观评估分析 - 12 -](#_Toc488500608)

[3.4.2 专家评估分析 - 17 -](#_Toc488500609)

[3.4.3 客观评估分析 - 20 -](#_Toc488500610)

[3.5 结论 - 26 -](#_Toc488500611)

[第4章 - 27 -](#_Toc488500612)

[第5章 - 28 -](#_Toc488500613)

[参考文献 - 29 -](#_Toc488500614)

[致 谢 - 30 -](#_Toc488500615)

# 引 言

# 相关研究

# 

本文研究基于搜索经验的查询推荐方法。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帮助用户记录搜索过程，并对复杂搜索任务进行管理，以及对复杂搜索过程进行可视化。本章研究时间树记录的用户的搜索过程中是否蕴含了搜索经验，为研究基于搜索经验的查询推荐方法提供理论基础。

## 时间树与用户的搜索经验

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树的节点，这些节点则被组织成为一种时序溯源的树型结构（Relative Chronological Source-tracking Tree ，*RCST*）。时间树理论中，与*RCST*结构相关的定义如下：

**【定义3.1】***RCST*节点：*RCST*节点*N*为二元组，。其中*G*为*RCST*节点类型。*C*为*RCST*节点内容。*T*表示*RCST*节点产生的时间。

下面给出*RCST*节点类型与*RCST*节点内容的定义。

**【定义3.2】***RCST*节点类型：*RCST*节点类型*G*为RCST节点的类型标识值，，若*G*的取值为查询类型，则称该*RCST*节点是一个查询节点，若*G*的取值为点击类型，则称该*RCST*节点是一个点击节点。一个*RCST*节点为查询节点表示该*RCST*节点是由用户的查询行为产生的，同理，一个*RCST*节点为点击节点表示该*RCST*节点是由用户的点击行为产生的。

**【定义3.3】***RCST*节点内容：*RCST*节点内容*C*是一段文本。*C*的值与*RCST*节点的节点类型有关，如果该*RCST*节点类型*G*为查询类型，则*C*表示该查询的查询词；如果该*RCST*节点类型为点击类型，则*C*表示该点击内容的标题。

**【定义3.4】***RCST*节点关系：*RCST*节点间关系*R(S,T)*为一个标识位，表示*S*与*T*具有兄弟关系或者父子关系。其中*S*与*T*是*RCST*节点，且*S*≠*T*。

在*RCST*节点关系中，*R(S,T)*并不具备广义上的关系的大部分性质，下面给出对*RCST*节点关系*R(S,T)*的具体解释：

1. *R(S,T)*可以有三种取值：*R(S,T)*是兄弟关系、*R(S,T)*是父子关系及*R(S,T)*是其他关系。如果*R(S,T)*为兄弟关系，则表示*S*是*T*的兄节点；如果*R(S,T)*为父子关系，则表示*S*是*T*的父节点；如果*R(S,T)*为其他关系，则表示*S*既不是*T*的兄节点也不是*T*的父节点。

时间树作为一种树形结构满足树形结构的一切性质。时间树的*RCST*结构从广义的树形结构中继承了父节点的定义，但广义的树形结构中并没有兄节点的概念，是*RCST*结构中特有的概念，下面针对时间树的*RCST*结构提出兄节点的定义。

**【定义3.4.1】**兄节点：*RCST*节点*A*是*RCST*节点*B*的兄节点，当且仅当*RCST*节点*A*与*RCST*节点*B*具有共同的父节点并且*RCST*节点*A*在时间树中的垂直位置高于*RCST*节点*B*。

时间树的*RCST*结构按照时间顺序自左向右，自上而下组织而成，根据时间树的*RCST*结构组织过程以及父节点及兄节点的定义可以得出如下结论：

**【结论3.4.1】**一个*RCST*节点*B*的兄节点*A*产生的时间*TA*一定早于*RCST*节点*B*产生的时间*TB*。

**【结论3.4.2】一**个*RCST*节点*D*的父节点*C*产生的时间*TC*一定早于*RCST*节点*D*产生的时间*TD*。

对于*RCST*节点*A*、*B*、*C*，若存在兄弟关系或父子关系*R(A,B)*及*R(B,C)*，则根据结论3.4.1与结论3.4.2，*RCST*节点A的产生时间*TA*早于*RCST*节点*B*的产生时间*TB*，并且*RCST*节点*B*的产生时间*TB*早于*RCST*节点*C*的产生时间*TC*。 因此*TA*一定早于*TC*。据此可以得出以下推论：

**【推论3.4.1】***RCST*兄弟关系与父子关系中，节点产生的时间*T*具有传递性。∀*RCST*节点*A*、*B*、*C*，若存在兄弟关系或父子关系*R(A,B)*及*R(B,C)*，则*RCST*节点*A*的产生时间*TA*一定早于节点*C*的产生时间*TC*。

1. *R(S,T)*具有非对称性，即*R(S,T)*不恒等于*R(T,S)*。如果*R(S,T)*为其他关系，则*R(T,S)*关系不能确定，可能为其他关系，也可能为兄弟关系或父子关系；如果*R(S,T)*为兄弟关系或父子关系，则*R(T,S)*一定为其他关系。
2. 对于父子关系，*R(S,T)*具有非传递性，即∀*RCST*节点*A*、*B*、*C*，若存在*RCST*关系*R(A,B)*及*R(B,C)*，则一定不存在*RCST*关系*R(A,C)*。
3. 对于兄弟关系，*R(S,T)*具有传递性，即∀*RCST*节点*A*、*B*、*C*，若存在*RCST*关系*R(A,B)*及*R(B,C)*，则一定存在*RCST*关系*R(A,C)*。
4. 对于兄弟关系与父子关系，*RCST*节点的产生时间*T*的先后关系具有传递性。既若*R(A,B)*与*R(B,C)*为兄弟关系或父子关系，则*RCST*节点*A*的产生时间一定早于*RCST*节点*C*的产生时间。

**【定义3.5】**时间树的*RCST*结构：时间树的*RCST*结构*T*为二元组，

。其中为*RCST*节点集，是*RCST*节点的集合，为*RCST*节点关系集，是*RCST*节点关系的集合。

时间树*RCST*结构的*RCST*节点关系集中包含了所有具有父子关系或者具有兄弟关系的节点对。基于*RCST*结构中的*RCST*节点关系，时间树能够帮助用户理解查询与点击之间的关系:

1. 用户可以通过时间树理解查询与查询，查询与点击之间的相对时间关系：

时间树上的节点与用户的搜索过程中产生的查询及点击行为一一对应，即时间树上的每个*RCST*节点代表用户搜索过程中的一次点击或者查询，并且用户搜索过程中的每一次点击与查询都被表示为时间树上的一个*RCST*节点。对于时间树上的任意两个RCST节点*A*、*B*，一定存在。

若*R(A,B)*不是其他关系，则依据结论3.4.1以及3.4.2， *RCST*节点*A*的产生时间*TA*一定早于RCST节点*B*产生的时间*TB*，确定了*RCST*节点*A*与*RCST*节点*B*的相对时间关系。

若*R(A,B)*是其他关系，并且在*RCST*关系集*RC*中能够找到一系列兄弟关系或父子关系{*R(A,C1)*,*R(A,C2)*,…,*R(Cn-1,Cn)*,*R(Cn,B)*}，则根据推论3.4.3，*RCST*兄弟关系及父子关系中的节点产生时间先后具有传递性，可以得到*RCST*节点*A*的产生时间*TA*早于*RCST*节点*B*的产生时间*TB*，确定了*RCST*节点*A*与*RCST*节点*B*的相对时间关系。

综上所述，时间树的*RCST*结构能够确定查询与查询，查询与点击之间的相对时间关系。

1. 用户可以通过时间树理解查询节点的来源：

时间树作为一种树形结构，除根节点外的所有节点都具有父节点。因此如果时间树上的一个查询节点*A*不是时间树根节点，一定存在父子关系。用户可以依据该关系理解查询节点的来源。

若*RCST*节点*X*的节点类型*GX*为点击类型，则*RCST*节点*A*的查询词*CA*来自于以*RCST*节点*X*的节点内容*CX*为标题的文档中。

若*RCST*节点*X*的节点类型*GX*为查询类型，则*RCST*节点*A*的查询词*CA*是对*RCST*节点*X*的查询词*CX*的一次查询重构。若不是查询重构，*RCST*节点*A*的查询词*CA*可能是来自任务要求或者用户自身的知识。

1. 用户可以根据时间树的*RCST*结构对复杂搜索任务进行子任务的划分。

对于时间树*RCST*节点集*NC*中的任一*RCST*节点*A*，至多有一个*RCST*节点使得父子关系，其中*RC*为*RCST*节点关系集。若存在*RCST*节点关系为父子关系，那么任意的*RCST*节点关系都不是父子关系。

但*RCST*节点可以有若干的子节点。即对于*RCST*节点集*NC*的子集*NCS*{*B,S1,S2,…,Sn*}，关系,,…,可以均为父子关系。

以上述子集*NCS*为例，时间树上的*RCST*节点*B*，在与*RCST*节点*S1,…,Sn*产生更多父子关系的同时也会使*RCST*节点*S1,…,Sn*两两之间产生更多的兄弟关系，当这些兄弟关系达到一定数量时，*RCST*节点*B*与它的兄弟节点间的视觉距离就会变大。这种视觉距离直观地将时间树分为几个部分，用户可以根据这种视觉距离进行子任务的划分。

时间树的*RCST*结构对用户理解查询与点击之间关系的帮助效果如图3.1所示。



图3.1

Fig. 3.1

1. *RCST*节点1～5具有父子关系*R(1,2),R(1,4),R(1,5),R(2,3)*，以及兄弟关系*R(5,4),R(5,2),R(4,2)*，根据结论3.4.1、结论3.4.2以及推论3.4.1，*RCST*节点1～5的产生时间先后顺序为1-5-4-2-3。
2. *RCST*节点1～4具有父子关系*R(4,1)*及*R(3,2)*，查询节点1的父节点为点击节点4，查询节点2的父节点为点击节点3，查询节点1来自用户点击的文章，查询节点2来自查询重构或题目要求及用户的自身的知识。
3. 由于以*RCST*节点1和2为根节点的子树中兄弟关系的增多，*RCST*节点1和2的视觉距离增大，用户可以直观清晰地进行子任务的划分。

时间树通过组织*RCST*结构，在*RCST*结构中记录*RCST*节点以及*RCST*节点关系记录用户的搜索过程。由此引出的一个问题是，时间树是否同时记录了用户的搜索经验以及基于这些搜索经验能否进行查询推荐。本章通过建立用户搜索经验模型，设计用户复杂搜索过程验证实验，并对实验结果进行分析，研究这一问题的第一部分，即时间树是否记录了用户的搜索经验。

为了回答时间树中是否蕴含用户搜索经验的问题，首先需要对用户的搜索经验进行模型化的描述。

## 用户搜索经验模型及用户搜索经验一致性模型

Habermas等人在2000年提出，人们在对过去事件的回顾过程中传达出用户对于过去事件的时间，因果，主题三方面的经验，这三方面的经验在用户对过去事件回顾过程中的时间一致性，因果一致性，主题一致性三种一致性中体现[1]。本文以此为理论依据，定义用户搜索经验模型如下：

**【定义3.6】**用户搜索经验模型：用户搜索经验模型*E*为一个三元组，，其中*TE*为用户时间经验，*CE*为用户因果经验，*THE*为用户主题经验。

1. 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。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。在进行搜索的过程中，用户的查询和点击会按照时间顺序先后进行，而不可能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时间先后顺序。
2. 因果经验为用户所进行查询的查询来源。用户进行的某个查询，一定是出于某种原因，例如查询词来自点击结果，或者查询词来自用户当前进行任务的任务需求，或者查询词来自用户的现有经验。
3. 主题经验为用户对复杂搜索任务所进行的子任务划分。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。这些子任务提供了一种复杂搜索任务的清晰结构展现，反映了用户对复杂搜索任务的规划。

时间经验，因果经验，主题经验共同组成了用户在复杂搜索过程中的搜索经验模型，但三种搜索经验无法被直观观测或直接衡量。为了解决这一问题，本文采用Habermas等人提出的搜索经验评价方法对搜索经验进行评价，提出搜索经验一致性模型，定义如下：

**【定义3.7】**用户搜索经验一致性模型：用户搜索经验一致性模型*C*为一个三元组，，其中*TC*为用户在对搜索过程回顾过程中的时间一致性，*CC*为用户在对搜索过程回顾过程中的因果一致性，*TH*为用户用户在对搜索过程回顾过程中的主题一致性。

1. 时间一致性指用户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。
2. 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。
3. 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。

如果时间树能够有效记录用户搜素经验*E<TE, CE, THE>*，那么用户在依据时间树进行搜索过程回顾时能够获得较高的时间一致性，因果一致性以及主题一致性，在不使用时间树进行搜索过程回顾时，在用户搜索经验一致性模型*C*上应当有较差表现。

基于上述分析，本章从用户对复杂任务搜索过程的回顾的角度设计实验，依据用户在对复杂搜索过程进行回顾时在时间一致性，因果一致性以及主题一致性上的表现来验证时间树对用户搜索经验的有效记录。

## 实验设计

本节实验的主要目的是以用户对复杂搜索过程进行回顾为载体，基于用户搜索经验一致性模型，通过分别考察用户使用时间树与不使用时间树，考察时间树能否更有效地维护搜索经验的一致性。

### 复杂搜索任务设计

为了针对上述目的进行实验的设计，首先需要设计搜索场景以使用户的搜索过程满足复杂搜索的定义。本文采用Singer等提出的复杂搜索任务的定义[2]来定义复杂搜索任务——“复杂搜索任务需要包括以下行为中的一种或几种：汇总，探索和组合。”以此为理论基础，本节进一步参考White等提出的用户可能需要在其中进行复杂搜索的三种典型场景[3]：1）系统对信息的索引不够充分；2）搜索任务本身需要浏览及探索；3）用户对搜索任务领域的知识匮乏以至难以对查询进行组织或定位信息领域。以这三种复杂搜索的典型场景为原则，设计复杂搜索任务，使实验参与者进行的搜索过程满足复杂搜索场景。

首先为了满足White等提出的复杂搜索的第一种典型场景——系统对信息的索引不够充分，实验采用学术搜索引擎进行。在学术搜索引擎中，受到版权及索引复杂度的限制，文章常常不会被全文索引，而仅仅会对标题及摘要进行索引，因此符合“系统对信息的索引不够充分”的场景需求。

为了满足White等提出的复杂搜索的后两种典型场景，实验共设计了两类共4个任务。

前两个搜索任务被称为学习型任务，采用食盐参与者的母语汉语进行布置。这两个搜索任务的话题都围绕实验参与者不熟悉但在日常生活中能够略微接触到的领域，因此实验参与者需要围绕任务话题进行浏览探索，这符合White等提出的第一种复杂搜索典型场景，是那些“需要浏览和探索”的任务。

后两个搜索任务被称为试探型任务，采用实验参与者的非母语英语进行布置。参与者不被直接告知搜索任务的具体描述，而是首先被要求阅读一篇英文摘要，在两天以后，参与者被告知需要围绕之前阅读的材料凭记忆进行搜索。由于人类记忆曲线的作用以及非母语环境下参与者对文章内容理解不够充分，实验参与者难以对查询进行组织，只能凭借记忆进行试探性查询探索，这符合White等提出的第二种复杂搜索典型场景，是那些用户“用户对搜索任务领域的知识匮乏以至难以对查询进行组织或定位信息领域”的任务。

对于学习型任务，参与者被要求根据任务要求完成搜索报告；对于试探型任务，参与者被要求完成一份与阅读材料相关的文章标题清单。两个学习型任务的任务描述及两个试探型任务的英文摘要出处如下：

* **任务一**：搜索相关论文，了解化疗中的常用药物，药物机理，使用方法，适用情况及副作用等，尤其针对癌症治疗，了解化疗药物的治疗手段，组合方式等。(不仅仅要了解支持化学疗法的药物，还要求了解对化学疗法进行辅助的药物，如缓解化学疗法副作用的药物等。)
* **任务二**：完成一篇报告，内容包括PM2.5 的概念、来源，PM2.5 的危害以及危害原因。找到中国 PM2.5污染最为严重的几个城市，并分析其形成原因。研究降低 PM2.5 的方法，并分析其原理以及其实际实施可能性(包括但不局限于利弊)。
* **任务三**：S. Shunmuga Krishnan, Ramesh K. Sitaraman: Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs. IEEE/ACM Trans. Netw. 21(6): 2001-2014 (2013)
* **任务四**：Hanqiang Cheng, Yu-Li Liang, Xinyu Xing, Xue Liu, Richard Han, Qin Lv, Shivakant Mishra: Efficient misbehaving user detection in online video chat services. WSDM 2012: 23-32

### 实验过程设计

本实验通过分别考察用户在对复杂搜索过程回顾中的时间一致性，因果一致性和主题一致性来综合考察时间树在用户对复杂搜素过程回顾中搜索经验一致性模型*E*上的表现，从而考察时间树能否更有效地维护搜索经验的一致性。

为了衡量用户使用时间树对复杂搜索过程回顾过程中的三种一致性，本实验采用按照时间顺序线性地记录查询和点击的方法（Timeline SearchLog，以下简称SearchLog）与时间树进行对比。SearchLog形式如图3.2所示。



图3.2 SearchLog示意图

Fig. 3.2

与普通浏览器日志类似，SearchLog按照用户进行查询点击的顺序记录查询与点击行为。其中查询以查询关键词的形式进行记录与展示，点击以点击链接标题加“----”前缀的形式进行记录与展示。

参与者被要求完成3.3.1节中所述的4个复杂搜索任务。每一位参与者分别从两类任务中被随机分配一个任务使用时间树进行搜索过程的记录与回顾，另一个任务则使用SearchLog进行搜索过程的记录与回顾。并且最终保证每个任务分别有4位参与者使用时间树和SearchLog进行搜索过程的记录与回顾。

由于实验所设计任务的复杂性，搜索难以一次性完成，并且为使实验场景满足Singer等提出的复杂搜索任务的定义，使实验参与者能够对实验任务进行深入的探索，本实验分为4个阶段进行，实验过程的设计如下：

1. 第一阶段：第一阶段为准备及搜索阶段，由于学习型任务与试探型任务在任务形式上的不同，第一阶段的设计有所不同。

学习型任务的第一阶段仅有搜索一部分，分3次进行，每次实验参与者搜索全部2个学习型任务，每个任务每次搜索时间限制为30分钟。

在复杂搜索过程中，由于搜索任务的复杂性，搜索者常常需要将复杂搜索任务分为多次完成，因此在复杂搜索过程中常常会发生搜索任务的中断与恢复。在搜索任务从中断状态恢复时，搜索者需要回忆自己已经从之前的搜索中获取到的信息，在这一过程中，蕴含了搜索者的回顾过程。因此，本实验利用这一特性，设计每次搜索的次间有1天的时间间隔，以探究在实验参与者将搜索任务从中断状态恢复时，时间树对参与者的帮助情况。

试探型任务的的第一阶段分为准备部分与搜索部分。

准备部分实验参与者被要求阅读2个试探型任务的英文阅读材料。非母语的阅读材料将大大增加用户组织查询的困难度，因此准备阶段不对实验参与者的阅读进行限时，以防止实验参与者在搜索阶段无词可搜的情况。

搜索部分在准备部分完成两天后开始，分2次进行，每次实验参与者搜索全部2个试探型任务，每个任务每次搜索时间限制为30分钟。出于与学习型任务一样的理由，次间有1天的时间间隔。

1. 第二阶段：第二阶段为报告阶段，在实验开始的一周后进行，时间为2天。参与者被要求完成报告，时间不限。

对于学习型任务，实验参与者需要在报告中汇报自己的搜索成果，搜索成果应当覆盖搜索任务中涉及的所有内容。

对于试探型任务，实验参与者需要在报告中完成一份文章清单，清单中包括参与者认为与所给阅读材料内容相关的文章标题。

1. 第三阶段：第三阶段为搜索过程调查及小组讨论阶段，在第二阶段结束后进行，持续时间为1小时。在第三阶段中，组织小组讨论，要求参与者探讨时间树使用中的主观感受并且填写主观调查问卷，问卷包括11个问题，具体的问卷内容及结果分析将在3.4节中介绍。
2. 第四阶段：第四阶段为搜索过程回顾及回顾过程调查阶段，在第三阶段结束一周后进行。在这一阶段实验参与者将使用时间树或SearchLog对第一阶段中进行的搜索任务进行回顾，并进行讲解。

为了保证实验参与者在对搜索任务进行回顾的过程中尽量依靠时间树以及SearchLog进行回忆，避免参与者在有所准备的情况下本能地对搜索过程进行记忆，增加实验的干扰因素，在整个实验进行到第四阶段以前，实验参与者都不被告知需要进行搜索过程的回顾与讲解。

参与者首先将被允许根据时间树或SearchLog对自己进行搜索过程进行回忆，回忆完成后将依据时间树或SearchLog进行搜索过程的讲解，讲解过程中将有两名专家对其讲解过程进行评分。在参与者的讲解过程中，专家可以随时针对参与者的讲解提出问题。在讲解结束后，参与者被要求对搜索任务进行子任务划分，并需要根据时间树或SearchLog对随机生成的10个查询关键词给出查询原因。

在上述回顾与讲解过程完成后，针对回顾的过程填写调查问卷，问卷包括3个问题，具体问卷内容及结果分析将在3.4节中介绍。

为了最大限度减少实验的干扰因素，实验需在可控环境下进行。实验环境控制原则如下：

1. 参与者只可以使用百度学术进行搜索。
2. 参与者只可以进行搜索任务范围内的搜索。
3. 参与者之间不允许进行任务相关话题的讨论。
4. 参与者在实验时间以外，尽最大可能不对任务相关话题进行思考。

## 实验结果分析

本实验的第三阶段及第四阶段中，对实验参与者的搜索及回顾讲解过程进行了主观评估、专家评估及客观评估三种评估，本节将针对三种评估的结果分别进行分析。

### 主观评估分析

实验参与者在实验的第三阶段以及第四阶段分别完成了一份调查问卷，用于实验参与者在搜索过程中以及对搜索过程进行回顾的过程中主观感受的研究，本节分别对两次调查的结果进行分析，从用户感受的角度对时间树维护一致性的效果进行评估，称为主观评估。

1. 复杂搜索过程的主观评估：

第三阶段的调查是关于用户在使用时间树进行复杂搜素过程中的主观感受。搜索过程主观调查通过给出一系列断言并要求实验参与者对这些断言进行评分，以此来评估参与者在使用时间树进行搜索的过程中的主观感受。

复杂搜索过程主观调查分为3个部分共11个断言组成。第一部分为断言1～断言4，用于通过实验参与者的主观感受验证时间树的易用性；第二部分为断言5～断言7，用于考察实验参与者是否感受到在搜索过程中利用了时间树的RCST结构进行了搜索经验的提取；第三部分为断言8～断言11，用于验证实验参与者将搜索任务从中断状态恢复时，参与者能否感受到时间树对参与者提供了帮助。

搜索过程主观调查的3个部分共11个断言具体如下：

第一部分：

* 断言1：很容易学习如何使用时间树，在学习过程中不需要花费很多的时间和精力。
* 断言2：时间树是非常易于操作的。
* 断言3：使用时间树来管理搜索任务没有带来额外的负担。
* 断言4：相比使用时间树所带来的额外负担，使用时间树所带来的好处更多。

第二部分：

* 断言5：我曾经利用过时间树节点的相对位置来判断查询和点击的先后顺序。
* 断言6：我曾经利用过时间树节点的父子关系来回忆我为什么要搜索某个查询。
* 断言7：我曾经利用过时间树的树形结构来划分搜索子任务。

第三部分：

* 断言8：在每次间隔两天再继续学习型搜索任务时，与不使用时间树相比，使用时间树能帮助我更好地回忆起上次的搜索任务进展到哪里，以及下一步应该做什么。
* 断言9：时间树很直观地展示了我的学习型任务搜索过程。
* 断言10：在每次间隔两天再继续试探型搜索任务时，与不使用时间树相比，使用时间树能帮助我更好地回忆起上次的搜索任务进展到哪里，以及下一步应该做什么。
* 断言11：时间树很直观地展示了我的试探型任务的搜索过程。

下面针对搜过过程主观调查的每一部分分别进行结果分析：

搜索过程主观调查结果第一部分如表3.1所示。

表3.1 搜索过程主观调查结果第一部分

Table 3.1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 断言编号 | 断言 | 评分范围 | 最高评分 | 最低评分 | 平均评分 | 标准差 |
| 1 | 断言1 | 0-10 | 10 | 8 | 9.25 | 0.886 |
| 2 | 断言2 | 0-10 | 10 | 7 | 8.25 | 0.886 |
| 3 | 断言3 | 0-10 | 9 | 4 | 7.38 | 1.51 |
| 4 | 断言4 | 0-10 | 10 | 5 | 8.88 | 1.73 |

搜索过程主观调查第一部分的主要目的是考察参与者主观感受中时间树是否易用。

观察断言1和断言2的最高评分、最低评分、平均评分及评分标准差可以得出，参与者认为时间树是易于学习并且易于操作的。

断言3的最高评分分别为9分，平均评分为7.38，但是最低评分4分，并且标准差偏大，说明参与者在时间树在使用中带来的负担问题上，参与者的主观感受上有所不同。在小组讨论部分，参与者提出的使用时间树时遇到的问题解释了这一评分结果。参与者认为，使用时间树时带来的负担主要来自于搜索过程中对时间树节点的调整。时间树的RCST结构中很重要的一个性质是具有来源追踪能力，但在实验场景下，由于捕获参与者当前查询来源所使用的技术尚不成熟，造成偶尔需要参与者手动调整时间树节点的情况发生，参与者认为，这是时间树给他们带来负担的主要原因。因此可以认为，参与者的负担主要是由技术原因导致，而并非时间树的RCST结构本身，并且本断言获得的平均评分高于7分，可以认为，时间树给实验参与者带来的使用负担在可接受的范围内，断言4获得的评分也证明了这一点。

断言4是断言3的后续断言，为了考察在参与者认为时间树为其带来负担的前提下，是否认为使用时间树的优点大于其缺点。虽然断言4获得的最低评分偏低，标准差相较于其他问题偏大，但从断言4的最高评分及平均评分结果中仍可以看出，参与者倾向于认为使用时间树进行复杂搜索带来的好处要多于坏处。

综合断言1～断言4的参与者评分结果可以得出，在参与者的主观感受中，时间树具有较高的易用性。

搜索过程主观调查结果第二部分如表3.2所示。

表3.2 搜索过程主观调查结果第二部分

Table 3.2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 断言编号 | 断言 | 评分范围 | 最高评分 | 最低评分 | 平均评分 | 标准差 |
| 5 | 断言5 | 0/1 | n/a | n/a | 0.875 | n/a |
| 6 | 断言6 | 0/1 | n/a | n/a | 0.750 | n/a |
| 7 | 断言7 | 0/1 | n/a | n/a | 0.750 | n/a |

搜索过程主观调查第二部分的主要目的是考察实验参与者是否感受到在搜索过程中利用了时间树的RCST结构进行了搜索经验的提取。参与者为此部分断言的评分只有0和1两种，0代表不同意断言，1表示同意断言。因此，平均评分代表了同意该断言的参与者的比例。

断言5的平均评分为0.875，因此有百分之87.5%的参与者认为在使用时间树进行搜索的过程中利用过时间树节点的相对位置来判断查询和点击的先后顺序，即87.5%的参与者主观上认为自己利用时间树的RCST结构从时间树上获取到了时间经验。

断言6及断言7的平均评分虽然略低于断言5的平均评分，但它们仍然处于较高的范围内。说明多数参与者认为在使用时间树进行搜索的过程中利用过时间树的节点的父子关系来回忆为什么要搜索某个查询以及树形结构来划分搜索子任务，即多数参与者主观上认为自己利用时间树的RCST结构从时间树上获取到了因果经验以及主题经验。

断言5～断言7的评分结果可以说明，多数参与者主观上认为自己在使用时间树进行搜索的过程中从时间树上获取到了搜索经验模型中的时间经验，因果经验以及主题经验。搜索过程主观调查第三部分的结果进一步支持了这样的结论。

搜索过程主观调查结果第三部分如表3.3所示。

表3.3 搜索过程主观调查结果第一部分

Table 3.3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 断言编号 | 断言 | 评分范围 | 最高评分 | 最低评分 | 平均评分 | 标准差 |
| 8 | 断言8 | 0-10 | 10 | 8 | 9.38 | 0.518 |
| 9 | 断言9 | 0-10 | 10 | 7 | 9.25 | 0.886 |
| 10 | 断言10 | 0-10 | 10 | 8 | 9.25 | 0.886 |
| 11 | 断言11 | 0-10 | 10 | 8 | 9.00 | 0.756 |

搜索过程主观调查第三部分是对调查第二部分的场景细化，主要目的是考察实验参与者将搜索任务从中断状态恢复时，能否感受到时间树的RCST结构对参与者提供了帮助。

断言8与断言10分别针对学习型任务与试探型任务提出，综合观察它们所获得的最高评分、最低评分以及平均评分，可以得出参与者认为在将搜索任务从中断中恢复时，时间树能够帮助参与者进行任务进展情况的判断。但断言8所获得的评分标准差更小，说明对于学习型任务，参与者们达成这种认识的统一性更高。搜索经验模型所定义的三种经验中，时间经验与主题经验是判断任务进展情况的重要参考因素。由此可以得出，在参与者的主观感受中，参与者在将搜索任务从中断中恢复的过程中，参与者从时间树中提取了时间经验以及主题经验。

断言9与断言11分别针对学习型任务及试探型任务考察了时间树在展示搜索过程上带给实验参与者的主观感受。从获得的最高评分、最低评分、平均评分以及评分的标准差可以看出，参与者普遍认为时间树很直观地展示了学习型任务以及试探型任务的搜索过程。

1. 搜索过程回顾的主观评估

第四阶段的调查是关于参与者在使用时间树以及SearchLog进行复杂搜素过程的回顾中的主观感受，以搜索经验模型为理论依据分别针对时间树与SearchLog提出了3个问题，要求实验参与者依据自身感受对其进行范围在0-10范围内的整数评分，从实验参与者主观感受的角度，通过对比使用时间树与不使用时间树时参与者的主观评分，评估时间树在搜索过程回顾中对时间一致性，因果一致性以及主题一致性的维护效果。

搜索过程回顾主观调查中的3个问题如下：

* 问题1：使用该方法对我回忆起查询点击的先后顺序的帮助程度。
* 问题2：使用该方法对我回忆起为什么要搜索某个查询，或点击某个搜索结果的帮助程度。
* 问题3：使用该方法对我回忆起搜索过程中的子任务的帮助程度。

搜过过程回顾主观调查问题在时间树与SearchLog上获得的最高评分、最低评分、平均评分以及评分的标准差如表3.4所示。

表3.4 搜索过程回顾主观调查结果

Table 3.4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 问题 | 数据类别 | 时间树 | SearchLog | |
| 问题1 | 最高评分 | 10 | | 10 |
| 最低评分 | 7 | | 4 |
| 平均评分 | 8.50 | | 7.63 |
| 标准差 | 1.20 | | 1.85 |
| 问题2 | 最高评分 | 10 | | 8 |
| 最低评分 | 7 | | 0 |
| 平均评分 | 8.75 | | 6.63 |
| 标准差 | 0.886 | | 1.19 |
| 问题3 | 最高评分 | 10 | | 8 |
| 最低评分 | 7 | | 5 |
| 平均评分 | 8.88 | | 5.88 |
| 标准差 | 0.991 | | 2.85 |

对于问题1，时间树与SearchLog获得的最低评分分别为7与4，时间树明显高于SearchLog，并且时间树的评分标准差也小于SearchLog的评分标准差。但时间树与SearchLog获得的最高评分都是10分，并且时间树获得的平均评分仅略高于SearchLog获得的平均评分。由此可以得出，时间树与SearchLog在帮助参与者回忆查询点击的先后顺序上都带给参与者较优的主观感受。

对于问题2，时间树获得的最高评分为10，高于SearchLog获得的最高评分8。时间树获得的最低评分为7，而SearchLog获得了极低的最低评分0。并且时间树获得的平均评分高于SearchLog的平均评分，评分标准差低于SearchLog的评分标准差。由此可以得出，参与者认为时间树对回忆查询动机的帮助程度高于SearchLog对回忆查询动机的帮助程度。

对于问题3，时间树获得的最高评分与最低评分分别高于SearchLog获得的最高评分与最低评分。并且时间树获得的平均评分远高于SearchLog获得的平均评分，评分标准差远低于SearchLog的评分标准差。说明参与者普遍认为时间树对回忆起搜索过程中的子任务的帮助程度高于SearchLog对回忆起搜索过程中的子任务的帮助程度。

综合问题1～问题3的结果分析，在帮助参与者进行时间经验提取方面，参与者认为时间树与SearchLog都有较优的表现。在帮助参与者进行因果经验以及主题经验提取方面，参与者认为时间树的表现要优于时间树。

为了验证以上结论，本研究使用显著性水平α=0.05的配对T检验（Paired Student’s T-test）对搜索过程回顾主观调查结果进行了分析。

配对T检验用于在样本数量较小时，检验来自两个总体的配对样本的样本均值是否有显著性差异。当配对T检验对应的p-值（p-value）小于等于显著性水平α时，两样本来自总体的差异是显著的，当配对T检验对应的p-值大于显著性水平α时，两样本来自总体的差异是不显著的。

搜索过程回顾主观调查的检验结果如表3.5所示。

表3.5 搜索过程回顾主观调查T检验结果

Table 3.5

|  |  |
| --- | --- |
| 问题 | 配对T检验p-值 |
| 问题1 | **0.176** |
| 问题2 | 0.0103 |
| 问题3 | 0.0263 |

在帮助参与者回忆起查询点击的相互顺序方面，时间树与SearchLog获得的参与者评分的T检验p-值大于显著性水平0.05，即时间树与SearchLog获得的评分差异不是统计显著的。

在帮助参与者回忆查询动机及搜索过程中的子任务两方面，时间树与SearchLog获得的参与者评分的T检验p-值均小于显著性水平0.05，说明时间树与SearchLog在这两方面获得的评分差异是统计显著的。

综合上述分析，得出结论：在帮助参与者进行时间经验提取方面，参与者认为时间树与SearchLog都有较优的表现。在帮助参与者进行因果经验以及主题经验提取方面，参与者在使用时间树时的主观感受统计显著地优于使用SearchLog时的主观感受。

### 专家评估分析

在实验的第四阶段，实验参与者使用时间树或SearchLog对第一阶段中进行的搜索任务进行回顾，并进行讲解。两名博士研究生将作为专家，对实验参与者的回顾性讲解进行0-10范围内的整数评分。两名专家评分的平均值，将作为实验参与者回顾过程的专家评分。

专家评估部分将针对4个评价因子对参与者使用时间树以及SearchLog时所作的回顾性讲解作出质量评价。根据用户搜索经验一致性模型，专家评估的前三个评价因子分别为参与者在对搜索任务进行回顾过程中的时间一致性、因果一致性以及主题一致性。第四个评价因子为参与者的回顾过程带给专家的整体感受。

由于学习型任务与试探型任务在复杂搜索场景上差异，参与者在对两种类型任务进行回顾的过程中表现不同，带给专家的感受也有所不同，因此，下面分别针对学习型任务及试探型任务对专家评估结果进行分析。

1. 学习型任务的专家评估结果如表3.6所示：

表3.6 学习型任务专家评估结果表

Table 3.6

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 评价因子 | 时间树 | SearchLog |
| 1 | 时间一致性 | 7.88 | 8.13 |
| 2 | 因果一致性 | 8.19 | 6.94 |
| 3 | 主题一致性 | 8.25 | 6.88 |
| 4 | 整体感受 | 7.81 | 6.81 |

从结果中可以看出，时间一致性方面，参与者在对学习型任务的搜索过程进行回顾过程中，使用时间树获得的专家评分仅略低于使用SearchLog的专家评分。这与3.4.1中进行的主观评估结果一致，在学习型任务中，时间树与SearchLog在帮助参与者维护时间一致性方面都有较优的表现。

参与者使用时间树时在时间一致性方面的专家评分低于在因果一致性与主题一致性方面获得的专家评分；使用SearchLog时在时间一致性方面的专家评分高于在因果一致性与主题一致性方面获得的专家评分。说明在使用时间树对学习型任务的搜索过程进行回顾时，相对于时间一致性，参与者维护了更高的因果一致性与主题一致性，而在使用SearchLog对学习型任务的搜索过程进行回顾时，相对于因果一致性与主题一致性，参与者维护了更高的时间一致性。

在因果一致性与主题一致性方面，参与者在使用时间树时与使用SearchLog时相比均获得了更高的专家评分。说明专家认为，参与者在使用时间树时维护了更高的因果一致性与主题一致性。

使用时间树时参与者的回顾过程带给专家的整体感受高于使用SearchLog时参与者的回顾过程带给专家的整体感受，说明专家认为在使用时间树时，相比于使用SearchLog时，参与者能够更准确地回顾学习型任务的搜索过程并且进行更清晰的讲解。

为了更准确地验证上述结论，在对专家评估结果的分析中，同样采用了显著性水平α=0.05的配对T检验对专家评估结果进行了分析。配对T检验结果如表3.7所示。

表3.7 学习型任务专家评估配对T检验结果

Table 3.7

|  |  |  |
| --- | --- | --- |
| 编号 | 评价因子 | 配对T检验p-值 |
| 1 | 时间一致性 | 0.792 |
| 2 | 因果一致性 | 0.00107 |
| 3 | 主题一致性 | 0.000142 |
| 4 | 整体感受 | 0.00347 |

在时间一致性方面，参与者在使用时间树与SearchLog进行学习型任务的搜索过程回顾时获得的专家评分配对T检验p-值大于显著性水平0.05，说明从专家评估的角度，在帮助用户维护时间一致性方面，时间树与SearchLog的差异并不是统计显著的。

因果一致性、主题一致性以及整体感受三个方面，参与者在使用时间树与SearchLog进行学习型任务搜索过程回顾时获得的专家评分配对T检验p-值均小于显著性水平0.05，说明从专家评估的角度，在帮助参与者维护因果一致性、主题一致性以及更准确地回顾学习型任务的搜索过程并进行清晰的讲解方面，时间树与SearchLog的差异是统计显著的。

综合上述分析，在参与者对学习型任务进行回顾及讲解过程中，时间树与SearchLog都能够有效帮助用户维护时间一致性。与使用SearchLog时相比，在使用时间树时，参与者能够统计显著地获得更高的因果一致性与主题一致性。并且使用时间树，参与者能够更准确地回顾学习型任务的搜索过程并进行更清晰地讲解。

1. 试探型任务的专家评估结果如表3.8所示：

表3.8 试探型任务专家评估结果表

Table 3.8

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 评价因子 | 时间树 | SearchLog |
| 1 | 时间一致性 | 7.94 | 7.94 |
| 2 | 因果一致性 | 7.75 | 7.13 |
| 3 | 主题一致性 | 8.06 | 6.94 |
| 4 | 整体感受 | 7.69 | 6.88 |

时间一致性方面，参与者在对试探型任务的搜索过程进行回顾中，使用时间树与使用SearchLog获得了相同的专家评分。与学习型任务中的结果相似，说明在试探型任务中，时间树与SearchLog在帮助参与者维护时间一致性方面也都有较优的表现。

使用时间树时，时间一致性获得的专家评分略高于因果一致性获得的专家评分，略低于主题一致性的专家评分；使用SearchLog时，时间一致性获得的专家评分高于因果一致性与主题一致性获得的专家评分。说明在试探型任务中，专家认为SearchLog对时间一致性的维护依然同样高于对因果一致性与主题一致性的维护。但专家认为，参与者在使用时间树对试探型任务进行搜索过程回顾时，因果一致性低于时间一致性。

单独考察因果一致性，时间树在因果一致性方面获得的专家评分略高于SearchLog，但分差似乎并不明显。这一点将在下文假设检验中验证。

主题一致性方面，参与者使用时间树获得的专家评分高于使用SearchLog获得的专家评分。说明与学习型任务相同，从专家评估的角度，参与者在使用时间树进行试探型任务的搜索过程回顾时主题一致性高于参与者在使用SearchLog时的主题一致性。

从整体感受的专家评分可以看出，使用时间树时参与者的回顾过程带给专家的整体感受高于使用SearchLog时参与者的回顾过程带给专家的整体感受，说明专家认为在使用时间树时，相比于使用SearchLog时，参与者能够更准确地回顾试探型任务的搜索过程并且进行更清晰的讲解。

与学习型任务相同，为了更准确地验证上述结论，在对专家评估结果的分析中，采用显著性水平α=0.05的配对T检验对专家评估结果进行分析。配对T检验结果如表3.9所示。

表3.9 试探型任务专家评估配对T检验结果

Table 3.9

|  |  |  |
| --- | --- | --- |
| 编号 | 评价因子 | 配对T检验p-值 |
| 1 | 时间一致性 | **0.383** |
| 2 | 因果一致性 | **0.0555** |
| 3 | 主题一致性 | 0.00369 |
| 4 | 整体感受 | 0.00280 |

配对T检验结果验证了上文对试探型任务专家评估结果的分析：

在时间一致性与因果一致性两个方面，参与者在使用时间树与SearchLog进行试探型任务搜索过程回顾过程中获得的专家评分配对T检验p-值大于显著性水平0.05，说明在试探型任务中，从专家评估的角度，在帮助用户维护时间一致性与因果一致性方面，时间树与SearchLog的差异不是统计显著的。

而在主题一致性与整体感受两个方面的专家评分配对T检验p-值小于显著性水平0.05，说明试探型任务中，从专家评估的角度，在帮助参与者维护主题一致性以及更准确地回顾学习型任务的搜索过程并进行清晰的讲解方面，时间树与SearchLog的差异是统计显著的。

综合上述分析，在参与者对试探型任务进行回顾及讲解过程中，时间树与SearchLog都能够有效帮助用户维护时间一致性以及因果一致性。与使用SearchLog时相比，在使用时间树时，参与者能够统计显著地获得更高的主题一致性。并且使用时间树，参与者能够更准确地回顾试探型任务的搜索过程并进行更清晰地讲解。

### 客观评估分析

1. 查询点击行为分析：

在3.4.2专家评估中，分析结果表明，时间树在维护参与者复杂搜索过程回顾中的搜索经验一致性上的表现，在学习型任务以及试探型任务上有所不同，这可能是因为两种任务本身的不同导致了参与者行为上的不同，从而导致了参与者对搜索过程回顾时的表现不同。因此，为了考察学习型任务与试探型任务是否导致了实验参与者不同的搜索行为，本实验在客观评估中首先统计了参与者在对两类任务进行搜索过程中产生的查询行为及点击行为个数。在本统计中将参与者在同一个关键词查询结果中的翻页行为看作参与者提交了一次新查询，统计结果如表3.10所示。

表3.10 查询／点击行为统计表

Table 3.10

|  |  |  |  |
| --- | --- | --- | --- |
| 搜索行为类型 | 数据类别 | 学习型任务 | 试探型任务 |
| 查询行为 | 最大个数 |  |  |
| 最小个数 |  |  |
| 平均个数 | 58.4 | 46.4 |
| 标准差 |  |  |
| 点击行为 | 最大个数 |  |  |
| 最小个数 |  |  |
| 平均个数 | 54.9 | 32.2 |
| 标准差 |  |  |

观察查询行为及点击行为在两类任务中的最大个数、最小个数、平均个数及标准差，试探型任务中查询行为个数明显大于点击行为个数，而学习型任务中查询行为个数仅略大于点击行为个数。

为了更准确地验证查询行为个数与点击行为个数关系在两种类型任务中的不同，对参与者在两种类型任务中的查询行为个数及点击行为个数进行了显著性水平α=0.05的配对T检测，结果如表3.11所示。

表3.11 查询／点击行为配对T检测结果表

Table 3.11

|  |  |
| --- | --- |
| 任务类型 | 配对T检验p-值 |
| 学习型任务 | **0.123** |
| 试探型任务 | 3.82e-5 |

结果表明，参与者在试探型任务中的查询行为个数与点击行为个数的配对T检测p-值小于显著性水平0.05，差异是统计显著的。而学习型任务中的查询行为个数与点击行为个数的差异不是统计显著的。说明参与者在试探型任务中相对进行了更少的点击行为。

1. 因果一致性及主题一致性的行为分析：

在5.1及5.2节对主观评估及专家评估结果的分析中，参与者及专家都认为时间树在帮助参与者维护因果一致性与主题一致性方面相对SearchLog的优势是统计显著的。这引出了一个问题，是什么导致了两种系统在帮助参与者维护因果一致性与主题一致性方面具有这种统计显著的差异。为了解释这种显著性差异产生的原因，本研究对参与者在使用时间树及SearchLog时体现因果一致性及主题一致性的行为进行了分析。

在本实验的第四阶段参与者对搜索过程进行回顾及讲解阶段，参与者被要求对搜索任务进行了子任务划分，并且对随机抽取的10个查询关键词进行查询原因的解释。对于学习型任务及试探型任务，参与者使用时间树及SearchLog时，划分的子任务及给出明确查询原因的查询关键词的平均个数如表3.12所示。

表3.12 因果一致性及主题一致性维护行为表

Table 3.12

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 一致性 | 学习型任务 | | 试探型任务 | |
| 时间树 | SearchLog | 时间树 | SearchLog |
| 因果一致性：查询原因明确的查询关键词平均个数 | 8.06 | 8.13 | 8.44 | 8.00 |
| 主题一致性：子任务划分平均个数 | 4.00 | 3.88 | 2.50 | 2.25 |

结果显示，参与者对自己所进行的每一种任务，随机抽取的10个查询关键词中，能够给出明确查询原因的查询关键字个数以及划分的子任务平均个数几乎相同。

为了更为严谨地分析参与者使用使用时间树及SearchLog在各种任务上能够给出明确查询原因的查询关键字个数以及划分子任务平均个数的相似程度，这一部分分析采用了显著性水平α=0.05的F检验（F-test）与配对T检验两种假设检验方式对上述问题进行分析。

在之前的分析中已经介绍过配对T检验的作用。本分析同时使用了F检验的原因是…。F检验与配对T检验的结果如表3.13所示。

表3.13 因果一致性及主题一致性维护行为假设检验结果表

Table 3.13

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 一致性 | F检验p-值 | | 配对T检验p-值 | |
| 学习型 | 试探型 | 学习型 | 试探型 |
| 因果一致性：查询原因明确的查询关键词平均个数 | 0.978 | **0.0171** | 0.935 | 0.573 |
| 主题一致性：子任务划分平均个数 | 0.791 | 0.948 | 0.781 | 0.678 |

观察查询原因明确的查询关键词平均个数及子任务划分平均个数在时间树与SearchLog上的F检验p-值以及配对T检验p-值可以得出，参与者在各种任务上使用时间树以及SearchLog能够给出明确查询原因的查询关键字个数以及划分子任务平均个数部具有显著性差异。

1. 因果一致性及主题一致性的表现因子分析：

上文中对参与者能够明确查询原因的查询关键字个数以及划分的子任务平均个数的分析结果似乎与3.4.1节主观评估分析以及3.4.2节专家评估分析中的结果相矛盾。由此引出一个问题，是什么导致了参与者在使用时间树对搜索任务进行回顾时在因果一致性与主题一致性方面能够获得比使用SearchLog更优的主观评估及专家评估结果，为了回答这一问题，本研究从参与者对搜索任务的回顾过程中提取了反映参与者因果一致性以及主题一致性的正面表现因子以及负面表现因子。

提取出的因果一致性正面表现因子及负面表现因子如下：

* 因果一致性正面表现因子：
* 清晰的来源描述，例如参与者这样的描述：“我读到了一篇……所以我搜索了……”
* 因果一致性负面表现因子：
* 在来源描述中缺少确定性，例如参与者这样的描述：“应该是有一篇……里面说了……吧。”
* 缺少查询来源的描述。
* 在讲解中被问到关于查询来源的问题时，不能较好地回答。

提取出的主题一致性正面表现因子以及负面表现因子如下：

* 主题一致性正面表现因子：
* 对查询和点击清晰的主题描述。
* 主题一致性负面表现因子：
* 不清晰的主题描述，例如参与者的描述中带有“可能……”,“差不多……”,“发现了几篇相关文章……”（没有说明这些文章是关于什么的。）
* 在讲解过程中没能较好地回答主题相关的问题。

在提取出因果一致性以及主题一致性的正面表现因子及负面表现因子后，针对每个任务，本研究对每位实验参与者对搜索过程的讲解中各表现因子的出现次数进行了统计。正面因子出现次数与负面因子出现次数的关系如图3.3所示

负面因子出现次数

正面因子出现次数

图3.3 正／负面表现因子关系图

Fig. 3.3

图3.3a～图3.3d分别表示学习型任务和试探型任务中因果一致性正负面表现因子关系以及主题一致性正负面表现因子关系。其中横坐标为实验参与者在对搜索任务进行回顾后的回想过程中正面因子出现的次数，纵坐标为这一过程中负面因子出现的次数。观察图中节点的分布可以看出，正面因子出现次数增多时，负面因子出现次数会减少。基于这一观察，本研究计算了正负面表现因子出现次数的差值，观察是否是正负面表现因子导致了专家认为参与者在对搜索任务进行的回顾后讲解过程中对因果一致性以及主题一致性的维护表现的不同。

下面将分别针对学习型任务以及试探型任务对因果一致性及主题一致性的正负表现因子出现次数差值与专家评分的关系进行分析。

* 学习型任务

学习型任务的正负表现因子出现次数差与专家评分的关系如图3.3所示。

正负面表现因子出现次数差值

专家评分

图3.4 学习型任务正负面表现因子出现次数差值与专家评分关系图

Fig. 3.4

图中横坐标为参与者在对学习型任务进行回顾后讲解过程中正负面表现因子出现次数的差值，纵坐标为专家评分。从图中观察到，在学习型任务中，正负面表现因子的出现次数差值与专家评分呈线形相关。为了验证这一观察，本研究采用显著性水平为0.05的皮尔逊相关检测（Pearson’s correlation test）对实验参与者在对学习型任务的回顾后讲解过程中，正负面表现因子出现次数的差值与专家评分之间的相关性进行检测。检测结果如表3.14所示。

表3.14 学习型任务正负表现因子出现次数差值与专家评分皮尔逊相关检测结果

Table 3.14

|  |  |  |
| --- | --- | --- |
| 一致性 | 相关系数 | p-值 |
| 因果一致性 | 0.752 | 7.85e-4 |
| 主题一致性 | 0.525 | 0.0367 |

结果表明，在在学习型任务中，因果一致性与主题一致性的正负表现因子出现次数差值与专家评分的相关系数较高，并且p-值均小于0.05。

根据上述分析可以得出结论，学习型任务中，因果一致性与主题一致性的正负面表现因子出现次数差值都与专家评分具有统计显著的线性相关关系。

* 试探型任务

正负面表现因子出现次数差值

试探型任务的正负表现因子出现次数差与专家评分的关系如图3.5所示。

专家评分

图3.5 试探型任务正负面表现因子出现次数差值与专家评分关系图

Fig. 3.5

图中横坐标为参与者在对试探型任务进行回顾后讲解过程中正负面表现因子出现次数的差值，纵坐标为专家评分。与学习型任务中一样，观察图中节点分布，发现在试探型任务中，正负面表现因子的出现次数差值似乎也与专家评分呈线形相关。为了验证这一观察，同样采用显著性水平为0.05的皮尔逊相关检测对实验参与者在对试探型任务的回顾后讲解过程中，正负面表现因子出现次数的差值与专家评分之间的相关性进行检测。检测结果如表3.15所示。

表3.15 试探型任务正负表现因子出现次数差值与专家评分皮尔逊相关检测结果

Table 3.15

|  |  |  |
| --- | --- | --- |
| 一致性 | 相关系数 | p-值 |
| 因果一致性 | 0.481 | **0.0593** |
| 主题一致性 | 0.414 | **0.111** |

从表中可以看出，检测结果与学习型任务并不相同。

试探型任务中，因果一致性与主题一致性的正负表现因子出现次数差值与专家评分的相关系数分别为0.481与0.414，说明它们之间具有一定的线性相关性。但p-值分别为0.0593与0.111，均大于显著性水平0.05，说明这种相关性并不是显著的。

据此可以得出，试探型任务中，因果一致性与主题一致性的正负面表现因子出现次数差值与专家评分并不具有统计显著的线性相关关系。

综合上述分析，学习型任务中因果一致性与主题一致性的正负面表现因子出现次数差值与专家评分之间的线性相关关系是统计显著的，而试探型任务中它们之间的线性相关关系并不是统计显著的。

本章研究的目的是考察参与者在使用时间树进行复杂搜索过程的回顾过程中的一致性表现。在主观评估及专家评估中，时间树在维护因果一致性及主题一致性方面的表现均优于SearchLog，因此本研究在分析了正负表现因子出现次数差值与专家评分之间的关系的基础上，进一步考察了正负表现因子出现次数差在时间树与SearchLog之间的不同统计规律，分别针对学习型任务与试探型任务的因果一致性与主题一致性计算了参与者使用时间树与SearchLog时的正负表现因子出现次数的平均差值，并对参与者使用时间树与SearchLog时的差值进行了显著性水平α=0.05的配对T检验，结果如表3.16所示。

表3.16 时间树与SearchLog正负表现因子出现次数平均差值及配对T检验结果表

Table 3.16

|  |  |  |  |
| --- | --- | --- | --- |
| 一致性及任务类型 | 正负表现因子出现次数平均差值 | | 配对T检验p-值 |
| 时间树 | SearchLog |
| 因果一致性，学习型任务 | 1.13 | -2.25 | 0.0230 |
| 因果一致性，试探型任务 | 0.125 | -0.750 | **0.277** |
| 主题一致性，学习型任务 | 1.13 | -1.63 | 0.0106 |
| 主题一致性，试探型任务 | 0.500 | 0.125 | **0.528** |

结果表明，在学习型任务中，参与者使用时间树与SearchLog时的因果一致性与主题一致性正负表现因子出现次数差值具有显著性差异。同时从结果中也可以看出，使用时间树时因果一致性与主题一致性正负表现因子出现次数平均差值高于使用SearchLog时的平均差值。根据上述结果，本研究认为，参与者使用时间树进行搜索任务回顾后的讲解时，正面表现因子出现因子出现次数更多，负面表现因子出现次数更少，从而导致了参与者在主观评估及专家评估中得到更优的结果。

## 结论

# 

本研究在第三章中提出的搜索经验模型指出搜索经验包括时间经验，因果经验及主题经验。第三章中的实验验证了时间树中蕴含了优质的因果经验及主题经验。本章将研究如何从时间树中提取因果经验及主题经验，以及如何利用提取出的因果经验及主题经验，进行查询推荐。

## 搜索经验提取

### 基于查询-点击-查询序列识别的因果经验提取算法

### 基于时间树结构的子任务划分算法

### 基于子任务划分的主题经验提取算法

## 面向搜索经验的查询推荐及推荐结果可视化

### 子任务内部的查询推荐方法

### 跨子任务的查询推荐方法

### 面向查询起源的查询推荐结果可视化方法

# 

# 参考文献

# 致 谢