# Search trail的相关概念

V. Bush. As we may think. Atlantic Monthly, 3(2):37–46, 1945.

通过信息空间浏览人类足迹开拓者（trail blazers）走过的路径，可以隐含地表示被访问的项目之间的相似性和关联

P. Pirolli, J. Pitkow, and R. Rao. Silk from a sow’s ear: extracting usable structures from the web. In Proceedings of the ACM SIGCHI conference on Human Factors in Computing Systems, pages 118–125, 1996

基于随机模拟浏览行为（a stochastic simulation of browsing behavior）的信息轨迹驱动的辅助检索模型（information-scent motivated model）

S. Chakrabarti, B. Dom, P. Raghavan, S. Rajagopalan, D. Gibson, and J. Kleinberg. Automatic resource compilation by analyzing hyperlink structure and associated text. In Proceedings of the 7th International World Wide Web Conference (WWW-98), pages 65–74, 1998.

通过信息空间浏览人类轨迹开拓者走过的路径，可以隐含地表示被访问的项目之间的相似性和关联，这些可以被纳入到轨迹推荐系统

（在08年的文章中，同样是用轨迹来判断兴趣，但不是用来支持浏览，而是更针对目的相关性估计，规模也更大）

A. Wexelblat and P. Maes. Footprints: history-rich tools for information foraging. In Proceedings of the ACM SIGCHI conference on Human factors in computing systems, pages 270–277, 1999.

基于其他用户轨迹的域内导航系统

Pirolli, P., & Fu, W. T. (2003). Snif-act: A model of information foraging on the world wide web. In Proceedings of the ninth international conference on user modeling (pp. 45–54).

开发并验证了基于信息搜寻理论的网络导航行为的计算认知模型。每次搜索路径被遍历时，信息轨迹（information scent）会为用户如何利用他们的搜索策略为他们的搜索任务找到有用的和相关的信息留下重要的线索。

R. W. White and S. M. Drucker. Investigating behavioral variability in web search. In Proceedings of the 16th International Conference on World Wide Web (WWW-2006).

重建搜索轨迹的方法

Downey, D., Dumais, S., & Horvitz, E. (2007). Models of searching and browsing: Languages, studies, and applications. In Proceedings of the 20th international joint conference on artifical intelligence. In IJCAI’07 (pp. 2740–2747).

提出了一个搜索活动语言模型的框架，其中使用状态空间模式合并了会话发起、查询执行、SERP(搜索引擎结果页)单击、非SERP单击和会话结束（session initiation, query execution, SERP (Search Engine Result Page) click, non-SERP click, and session end）等活动

Ryen White, Mikhail Bilenko, and Silviu Cucerzan. 2007. Studying the Use of Popular Destinations to Enhance Web Search Interaction. In SIGIR. ACM, 159–166.

开发了一个流行的目的地工具，它显示的页面通常是类似查询的跟踪终点，该工具对探索性搜索任务有利

Wang, X. & Zhai, C. (2009). Beyond hyperlinks: organizing information footprints in search logs to support effective browsing. Proc. CIKM, 1237-1246.

允许用户横向导航到相关查询，纵向导航到不同特性的查询。具有预定义策略的模拟用户可以从这种映射中获益

生成/推荐搜索路径

\*使用户主动share搜索路径：提供一个接口，标记有用的信息/建立一个trail（增加了用户主动性之后，与之前系统推荐有什么区别）

类似于wiki，共同编辑一个trail

协作检索，没有直接交流情况下的协作检索，无法进行分工（可以互相推荐）

对哪种用户有用，对编辑的用户不要造成太大的麻烦，形成动机

评估有效性

可能会导致用户的懒惰，可能会有bias

使用户知情

隐形协作模式

用户需要一定的信息组织能力，贡献者本身也是一个搜索者，需要顺手点两下

反复搜索的东西，角色的转换

Search trail需要变成一个层级的结构

如何确定信息的权威性：用户做反馈，比较有争议性的东西需要汇集各种信息；搜索的过程中还是自组织

搜索引擎固有的简单性，只是一个简单的跳转，任务的复杂度

想出的具体的研究问题：变量：task（类型）

大众关心的话题

Citizen science：普通大众参与到科研里/协作搜索

搜索引擎比社会化问答社区更加精确

想一下具体的协作模式

# search trail的潜在信息价值

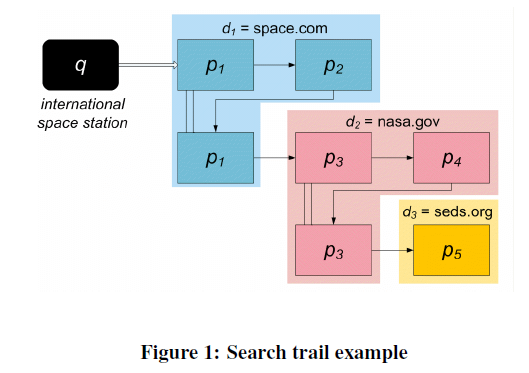
应该是最早的文献：

Mikhail Bilenko and Ryen White. 2008. Mining the Search Trails of Surfing Crowds. In WWW. ACM, 51–60.

使用“learning-to-rank”的方法，发现通过从完整的路径生成训练数据而不是仅通过serp级别的交互能够更好提高检索性能；

直接利用真实用户的浏览和搜索日志数据，从其大数据集中学习相关模型；执行两组实验验证从用户日志中确定相关资源的可行性：①排名准确性（网站层面）；②监督学习（supervised learning）算法

对路径上所有的页面进行推荐



Datasets：140 million search trails，The first dataset, HumanRanking, contains 33,150 queries randomly sampled based on their relative frequency from query logs of Windows Live Search。人工打分排序

The second dataset, UsageRanking, 10,000 queries, each query must have been entered into at least two different search engines. 按照网页的浏览量进行排序

很有参考价值

Hassan Awadallah, A., White, R. W., Pantel, P., Dumais, S. T., & Wang, Y.-M. (2014). Supporting complex search tasks. In Proceedings of the 23rd ACM international conference on conference on information and knowledge management . In CIKM ’14 (pp. 829–838).

提出了自动挖掘任务搜索日志并建立连接多个任务的关联图的方法，然后利用任务图来帮助新的搜索者探索新的搜索主题或处理多步骤搜索任务

定义复杂任务为①探索学习②某一主题③多步骤

首先，从queries里面辨别出任务，可以主题中的一个方面的任务，也可以是多步骤任务中的一个步骤①利用实体库检索出实体②多次关键字（collocations）③Prep介词，如(for, in, of等)。任何没有标记为实体、搭配或介词的都被称为term④patterns；

其次，将包含同一任务的所有query组合在一起，一种方法是定义查询相似度函数，并将其应用于所有对查询，然后对结果图进行聚类。或者，可以使用关于query的元数据(the refiner and pivot tags, the entity)来显著降低分组过程的计算成本；

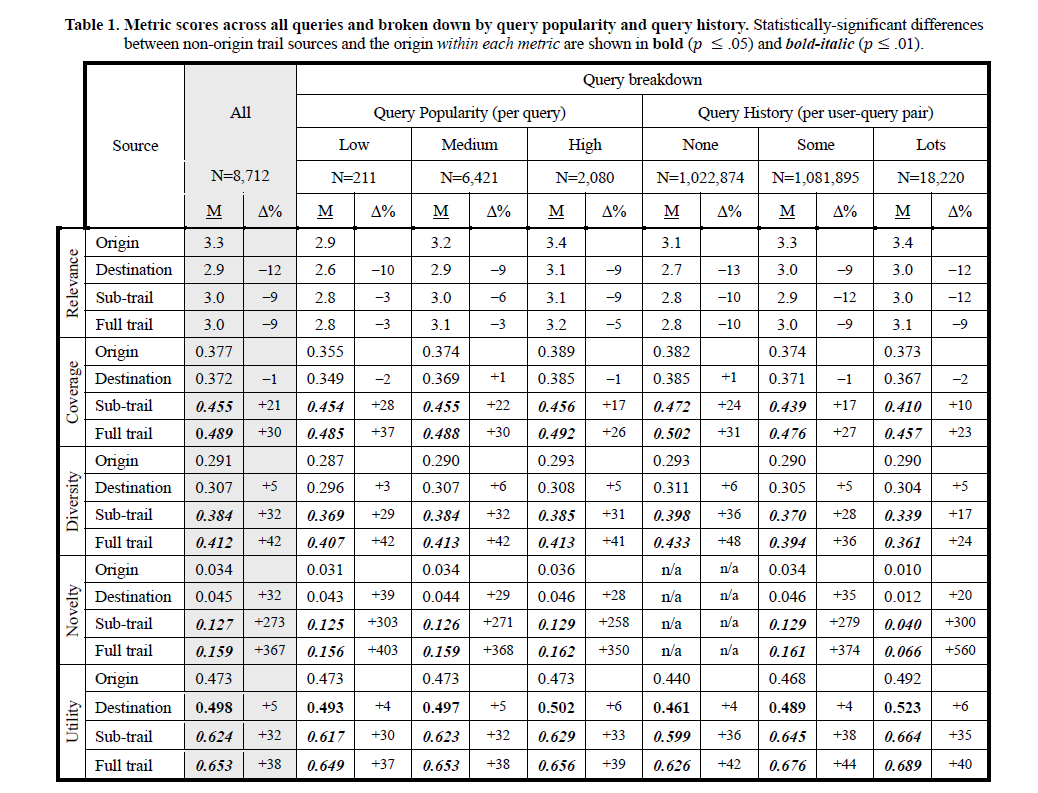
最后，绘制task graph，将所有知识连在一起

# 影响search trail质量的因素

Ryen White and Jeff Huang. 2010. Assessing the Scenic Route: Measuring the Value of Search Trails in Web Logs. In SIGIR. ACM, 587–594.

比较了完整的搜索路径和部分搜索路径的有用性，发现完整的搜索路径提供了更大的主题覆盖率、多样性和新奇性。





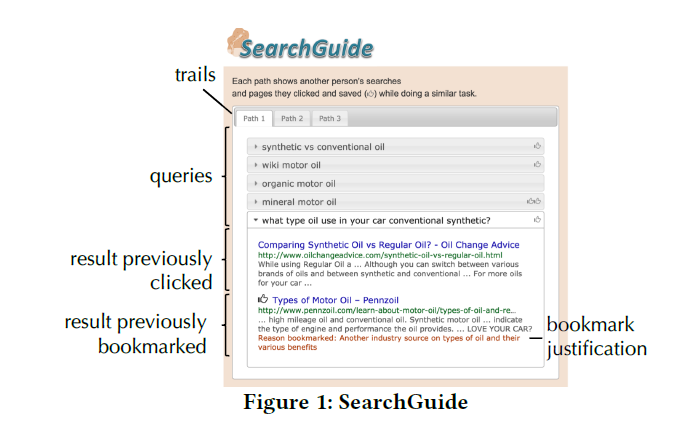
没有提到搜索路径在界面中是如何组织的

Xiaojun Yuan and Ryen White. 2012. Building the Trail Best Traveled: Effects of Domain Knowledge on Web Search Trailblazing. In SIGIR. ACM, 1795–1804.

发现由领域专家(相对于新手)创建的搜索路径包含更多的相关信息，更多事实(相对于主观)信息，以及从一般搜索到特定搜索的更合理（more logical）的过渡。

Robert Capra and Jaime Arguello. 2019. Using Trails to Support Users with Tasks of Varying Scope. In SIGIR. ACM, 977–980.

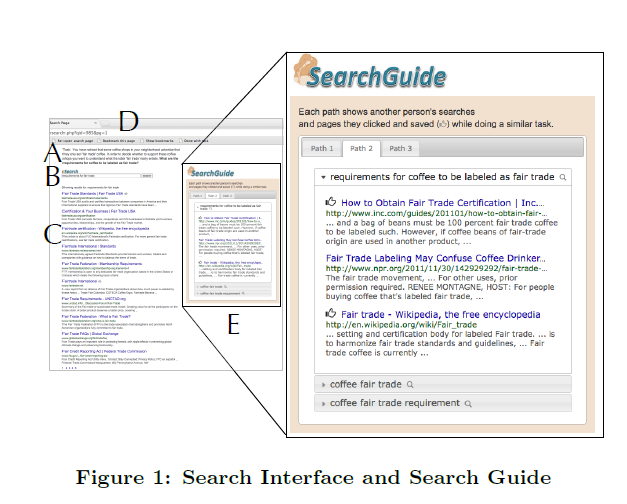
搜索任务的项目/维度对search trail的影响（维度更抽象，难度更大），搜索轨迹覆盖范围对search trail的影响（最好匹配，否则越大越好）



搜索界面上有：①任务描述②SG（见上图）③用户回答的文本框④结束任务按钮

Robert Capra, Jaime Arguello, Anita Crescenzi, and Emily Vardell. 2015. Differences in the Use of Search Assistance for Tasks of Varying Complexity. In SIGIR. ACM, 23–32.

从其他人的search trail中获益所面临的挑战



首先，对于**更复杂**的任务，与SG的交互作用更大。第二，任务的先天决定性帮助预测参与者是否从SG那里获得了书签。第三，与SG互动但没有获得书签的参与者比获得书签和没有互动的参与者感觉更少的系统支持。第四，通过对访谈的定性分析，我们发现，在任务复杂性的不同水平上，使用SG的动机和好处是不同的，对于最简单的(记忆)任务，当参与者使用SG时，主要是确认和验证信息。相比之下，对于更复杂的任务，他们使用它来寻找新的信息来源和新的搜索策略。第五，兴趣、先验知识和搜索经验不会影响SG的使用

# 为当前用户开发预测search trail的技术

Wheeldon, R. & Levene, M. (2003). The best trail algorithm for assisted navigation of web sites. Proc. LA-WEB, 166.

搜索路径通过树状图显示在浏览器界面上，帮助用户识别链接之间的联系

Adish Singla, Ryen White, and Jeff Huang. 2010. Studying Trail finding Algorithms for Enhanced Web Search. In SIGIR. ACM, 443–450.

通过识别除不同trail-finding算法的区别，以及如何在给定查询那种找出性能优于其他算法的search trail，展示了search trail的价值

Chathra Hendahewa and Chirag Shah. 2017. Evaluating User Search Trails in Exploratory Search Tasks. IP&M 53, 4 (2017), 905–922.

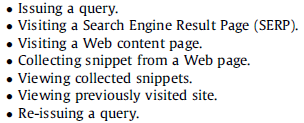
提出一种基于查询编辑距离(query edit-distance)的轨迹匹配算法

Search trail评估的两种方法：①信息空间覆盖组织：使用the DMOZ ODP-based categorization ( http://www.dmoz.org )计算search trail的深度和广度，如果目录中没有则手动标记。②不同查询路径的推荐顺序通过query和page的推荐顺序进行度量。

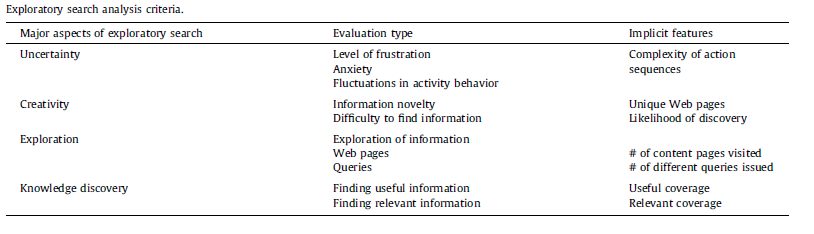
这篇文章没有写明，用户实验中如何分配search trail变量（随机/真实/推荐）

Hendahewa, C., & Shah, C. (2015). Implicit search feature based approach to assist users in exploratory search tasks. Information Processing & Management, 51 (5), 643–661.

将搜索过程分成几个阶段：



搜索过程的指标：



通过测量上述指标评估搜索路径

与之前的方法不同的是，还预测了用户在搜索过程中不同步骤的耗时，以此进行推荐，模拟实验表明，如果用户采纳了建议，早期的干预会改善搜索过程中的搜索性能。