

# 基于交互感知的探索式搜索中资源的推荐方法

孙海春 李欣

(中国人民公安大学信息技术与网络安全学院 北京 100038)

(中国人民公安大学安全防范与风险评估公安部重点实验室 北京 100076)

**摘要** 基于关键词匹配的信息搜索技术日渐成熟,人们期待搜索引擎更加智能,能够根据模糊需求启发用户找到正确的信息,探索式搜索由此出现。已有依据关键词之间的语义关联进行搜索推荐的探索式搜索方式,能够启发用户找到与当前查询相关的其他信息。然而,这种随机式的推荐方式仅适用于用户对不熟悉的领域进行初步了解的应用场景。针对有特定搜索意图的用户,如何有效利用用户提供的信息,让信息推荐以符合认知规律的方式进行,并且能够使计算机智能地理解用户的下一步信息需求的意图是研究的关键。提出一种基于交互感知的信息推荐方法,设定交互的基本规则,便于系统尽快确定模糊用户需求的搜索意图。示例展示这种方式能够对有效交互进行定量分析,并能有效减少用户与系统之间交互的步数。

**关键词** 探索式搜索,交互系统,有效交互,搜索意图

**中图法分类号** TP391 **文献标识码** A

## Resource Recommendation Method Based on Interactive Perception in Exploratory Search

SUN Hai-chun LI Xin

(College of Information Technology and Network Security, People's Public Security University of China, Beijing 100038, China)

(Key Laboratory of Safety Precautions and Risk Assessment, Ministry of Public Security,

People's Public Security University of China, Beijing 100076, China)

**Abstract** Information search technology based on keyword matching is becoming more and more mature. Search engines are expected to become more intelligent. As a kind of search engine which satisfies the information searching with fuzzy requirements, exploratory search comes up. Some existing methods which recommend information based on semantic relations among keywords can inspire users to find other information related to the current query. However, this kind of random recommendation can only achieve some application scenarios like users' understanding of unfamiliar areas. For users who have a concrete search intention, how to make computers understand their information requirement based on some prior information is the key issue. This paper proposed a method of information recommendation based on user interactive perception, which includes two basic rules of interaction to make system understand users' fuzzy requirements quickly. The example shows that it allows quantitative analysis of effective interactions. And, it can reduce interaction steps between users and systems.

**Keywords** Exploratory search, Interactive system, Effective interaction, Search intention

## 1 引言

探索式搜索可能仅仅由好奇心引发,它的搜索过程是机会性的、迭代的。例如,当用户在一个不熟悉的领域内进行调研时,他们想获取更多的知识来加深对该领域的认知,而不是有目的地去寻找某一类信息;或者,用户存在特定的搜索意图,但限于已有知识有限不知如何表达。此时,需要展现更多与当前搜索具有现实关联的信息给用户,而不能像现有搜索引擎那样,仅仅展示与搜索词匹配的信息给用户。

已有的探索式搜索方法大多依据词语之间的语义关联进行搜索推荐;一些成果通过挖掘用户搜索日志,提炼在一次搜索过程中共同出现或者具有逻辑关联的词语,并推荐给用户<sup>[3-4]</sup>。有些研究者通过本体中词语之间的语义含义或者客观存在的文本、视频等信息载体,挖掘出信息之间的现实语义

关联,并借助它们之间的关联大小推荐给用户<sup>[5-6]</sup>;或者基于社交网络的方式推荐相关信息<sup>[7]</sup>。

然而,与当前搜索关联的信息很多,哪些信息被优先推荐给用户决定着能否在最短时间内找到需要的信息。已有研究往往采用了最为简单的交互方式,仅仅依赖于用户输入的查询词,依据用户的日志记录、信息与查询词之间的语义关联程度来完成下一步查询关键词的推荐。实际上,对于探索式需求,用户的需求在用户浏览过程中逐步确定,与用户进行交互是获得当前用户意图的有效方式。一般来说,与用户进行交互是让用户逐渐确定自身需求的一种方式,也是一种让机器更好地理解用户需求并满足用户需求的有效途径。本文借助于用户的浏览行为,对搜索推荐的交互机制进行改进,便于用户更快找到潜在的需求信息。

如何有效利用用户提供的信息,让信息推荐以符合认知

规律的方式进行,并且能够使计算机智能地理解用户的下一步信息需求意图,是探索式搜索系统中交互设计所面临的挑战之一。本文定义并验证交互设计的两个基本原则,提出一种基于交互感知的资源推荐方法,使得用户在需求模糊的情况下,逐步明确自身需求,以最小的交互步数找到想要的信息。

## 2 相关工作

已有的探索式搜索大多依据词语之间的语义关联进行搜索关键词推荐。文献[5]将互联网络上的信息按照人工目录进行分类,利用超链接提取类之间的客观关联,并以此为基础提出了一种面向主题的探索式搜索方法。文献[6]利用信息之间的共现与文本结构、信息之间的链接结构,提取了信息之间的语义关联,并以此为基础完成关键词推荐。文献[3]基于语言建模和聚类来处理网络搜索日志,利用提炼出的搜索兴趣模式向用户推荐信息。文献[8]考虑用户的历史搜索与用户当前浏览的网页,以此来使得推荐体现用户个性化。以上方式关注于如何挖掘探索式搜索所基于的信息之间的语义关联,所给出的方法能够找到与当前搜索相关的信息。

已有很多工作集中在交互的可视化展示上,它们提供了很多可视化界面的实现方式,能够有效帮助用户快速理解所展示的信息。文献[1]给出了界面抽象元素和实体元素以及映射关系的形式化描述,提出了一种面向信息可视化的语义 Focus+Context 人机交互技术。文献[9]设计了一种查询预览控制可视化的新的检索方式,能够在使用预览的同时看到不同的相关文档。文献[10]基于 Web 语义开发了一种新的集成系统,同时体现了语义信息的数据模型、动态提取以及呈现方式,支持用户查找、发现和组合 Web 文本。然而,这些技术侧重于给用户提供更好的视觉体验,并未给出推荐方法,来判断哪些信息需要优先展现给用户。

综上所述,基于语义关联向用户推荐信息是一种有效的方式。然而,信息往往同时属于多个领域或者多个主题,仅仅基于语义关联进行信息推荐会将多个领域或主题的信息集中展现,显示杂乱且容易误导用户。利用相似用户的日志文件进行协同过滤的方法需要前期积累足够的用户日志,新用户或者用户进入新的领域搜索,都会因日志积累不足造成误判。融合信息之间的语义关联以及用户当前的搜索意图,是目前探索式搜索的研究方向之一。本文工作以信息之间的现实语义关联为基础,侧重使用用户本身的意愿进行信息推荐,将用户与系统之间的交互过程进行形式化,使得机器智能地逐渐明确用户的查询意图,从而找到需要的信息。

## 3 探索式交互的设计原则

与用户进行交互是用户使用户逐渐确定自身需求的一种方式。基于信息语义进行有效交互是正确理解用户意图的前提。

**定义 1(有效交互)** 交互对满足用户需求起正面导向作用,即为有效交互。

定义 1 给出了有效交互的定义。不同应用场景下,对满足用户需求起正面导向作用的理解不同。本文对定义 1 进一步细化,给出一种具体的有效交互的形式化描述方式,将有效交互细化为以下两个原则。

**原则 1** 如果 C 是 A-B 之间的有效交互,那么 C 到 B 的有效距离要小于 A 到 B 的有效距离。

已知起点和终点的情况下,交互的有效区域如图 1 所示。图 1(a)矩形部分标识了从资源 B 找资源 A 的有效交互区域;图 1(b)立方体部分标识了从资源 B 找资源 A 的有效交互区域。

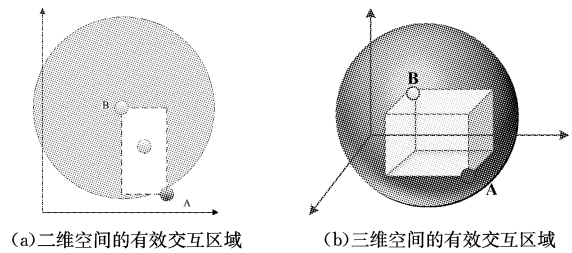


图 1 交互原则 1 满足的示意图

定义 2 给出了有效交互的形式化定义。

**定义 2** 假设当前查询词为  $k_0$ , 目标点为  $k$ , 系统给用户的推荐序列为:  $k_1, k_2, \dots, k_n$ 。如果用户选择  $k_i$ , 则认定  $k_i$  与  $k$  之间的距离最近, 即: 对于  $\forall k_j \in \{k_0, k_1, k_2, \dots, k_n\}, |k_i - k| \leq |k_j - k|$ 。其中,  $|k_i - k|$  表示  $k_i$  与  $k$  之间的语义距离。

**原则 2** 在与用户交互的过程中展示可区分度高的资源。

通常情况下符合用户表达需求的元素的集合是比较大的, 如何在该集合中优先选择与用户进行交互的  $n$  个 ( $n$  要在用户能够主观判断的范围之内, 通常情况下  $n \leq 10$ ) 元素, 是交互系统的关键之一。一般情况下, 推荐给用户的  $n$  个元素之间的离散程度越大, 在一步交互中, 系统能够确定的、不是用户搜索目标的元素越多, 那么, 系统和用户能够在最少步骤完成交互。本文中, 交互原则 2 被形式化定义为定义 3。

**定义 3** 资源集合  $K = \{k_1, k_2, \dots, k_l\}$  中,  $\{k_1', k_2', \dots, k_n'\}$  (其中,  $n \leq l$ ) 为  $K$  中可区分度最高的  $n$  个资源, 则以下条件满足:

$$\forall \{k_1'', k_2'', \dots, k_n''\} \subseteq \{k_1, k_2, \dots, k_l\}, \sum_{i,j \in \{1,2,\dots,n\}} |k_i'' - k_j''| \leq \sum_{i,j \in \{1,2,\dots,n\}} |k_i' - k_j'|$$

定义 3 将可区分度高的资源具体化为  $n$  个资源之间的语义距离最大。为了验证以上结论成立, 给出定理 1 及其证明。

**定理 1** 系统在集合中选取  $n$  个元素, 保证选取的  $n$  个元素相互之间的距离累加和最大时,  $n$  个元素之间的离散程度最大。

证明: 假定元素用  $m$  维向量表示, 选取的元素个数为  $n$ , 将方差作为元素之间离散程度的衡量指标。推导过程如下。

$n$  个元素之间的距离累加和为:

$$\sum_D \{x_1, x_2, \dots, x_n\} = \sum_{i,j \in \{1,2,\dots,n\}} \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{im} - x_{jm})^2}$$

$n$  个元素之间的离散程度 (Measures of Dispersion) 定义为:

$$\text{MOD}(\{x_1, x_2, \dots, x_n\}) = \frac{1}{m} \sum_{p=1}^m \frac{1}{n} \sum_{i=1}^n (x_{ip} - \bar{x}_p)^2$$

推导 MOD 的公式, 得:

$$\begin{aligned} \text{MOD}(\{x_1, x_2, \dots, x_n\}) &= \frac{1}{m} \sum_{p=1}^m \frac{1}{n} \sum_{i=1}^n (x_{ip} - \bar{x}_p)^2 \\ &= \frac{1}{nm} \sum_{p=1}^m \sum_{i=1}^n (x_{ip} - \bar{x}_p)^2 \\ &= \frac{1}{nm} \sum_{p=1}^m \sum_{i=1}^n \left( x_{ip} - \frac{x_{1p} + x_{2p} + \dots + x_{np}}{n} \right)^2 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{mn^3} \sum_{p=1}^m \sum_{i=1}^n [(x_{ip} - x_{1p}) + (x_{ip} - x_{2p}) + \dots + (x_{ip} - x_{np})]^2 \\
&= \frac{1}{mn^3} [(n-1) \sum_{i,j \in \{1,2,\dots,n\}} (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{im} - x_{jm})^2 + \sum_{i,j \in \{1,2,\dots,n\}} (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{im} - x_{jm})^2] \\
&= \frac{1}{mn^2} [\sum_{i,j \in \{1,2,\dots,n\}} (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{im} - x_{jm})^2]
\end{aligned}$$

令:

$$\sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{im} - x_{jm})^2} = d_{ij}$$

那么:

$$\sum_D \{x_1, x_2, \dots, x_n\} = \sum_{i,j \in \{1,2,\dots,n\}} d_{ij}$$

$$MOD(\{x_1, x_2, \dots, x_n\}) = \frac{1}{mn^2} \sum_{i,j \in \{1,2,\dots,n\}} d_{ij}^2$$

由于  $d_{ij} \geq 0$ , 因此  $\sum_D \{x_1, x_2, \dots, x_n\}$  与  $MOD(\{x_1, x_2, \dots, x_n\})$  正相关, 前者增大时后者也增大, 从而结论成立。

### 4 交互设计原则的有效验证示例

依据原则 1, 交互系统能够逐步缩小目标所在的范围, 最终找到目标。依据原则 2, 交互系统每次将符合搜索条件的目标集合中可区分度最大的  $n$  个元素展现给用户, 以帮助用户快速筛选掉大部分非相关元素, 从而能够在最少步骤内完成交互。示例如下。

例 1 假如查找的起始点 A 为(4,0), 目标点 B 为(4,2)。在不知道目标点的情况下, 系统随机选取了 4 个点推荐给用户, 分别为: (4,8), (-4,0), (12,0), (4,-8), 如图 2 所示。系统与用户的交互过程: 1) 选择 4 个点中与目标点最近的点, 用户选择点(4,8); 2) 起始点(4,0)与所选的点(4,8)哪个与目标点更接近? 用户选择(4,0)。因此, 根据交互原则 1, 系统得到如下不等式:

$$(x-4)^2 + (y-8)^2 \leq (x-12)^2 + (y-0)^2 \quad (1)$$

$$(x-4)^2 + (y-8)^2 \leq (x+4)^2 + (y-0)^2 \quad (2)$$

$$(x-4)^2 + (y-8)^2 \leq (x-4)^2 + (y+8)^2 \quad (3)$$

$$(x-4)^2 + (y-8)^2 \geq (x-4)^2 + (y-0)^2 \quad (4)$$

由以上 4 个不等式, 得到目标点的坐标点满足:

$$x-y \leq 4; x+y \geq 4; 0 \leq y \leq 4$$

由此, 系统可以推断出, 用户想要查找的目标点在图 2 中 3 条直线围成的三角形区域内。

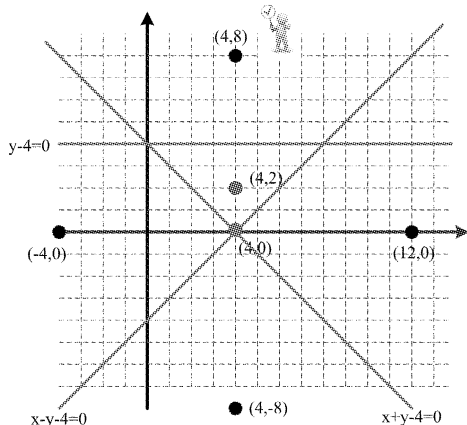


图 2 交互约束中原则 1 的应用示例(利用原则 1 缩小目标资源的范围)

例 2 上例中, 依据用户的选择, 选定了如图 3 所示的区域。除去起始点(4,0), 区域中共有 6 个元素: (4,1), (4,2), (3,2), (4,3), (5,3), (6,3)。设定  $n=3$ , 那么:

$$MAX \sum_d \{C_d^3\} = \sum_d \{(4,1), (3,2), (6,3)\} = 7.4$$

因此, 系统选择向用户展示的 3 个元素为: (4,1), (3,2), (6,3)。

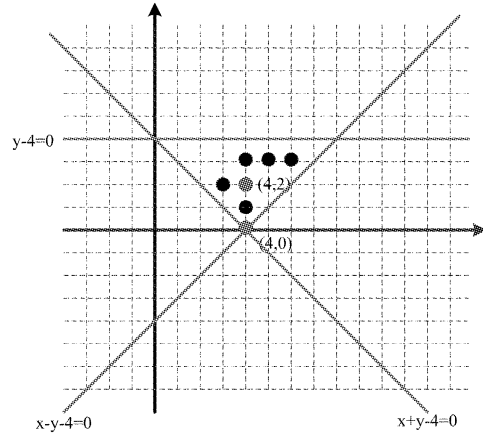
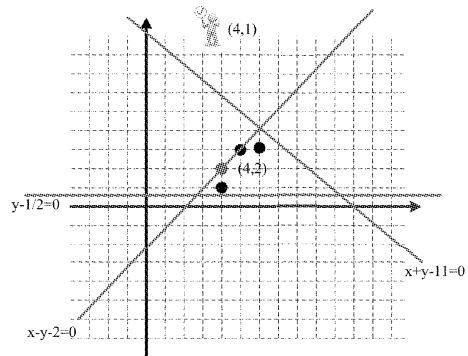
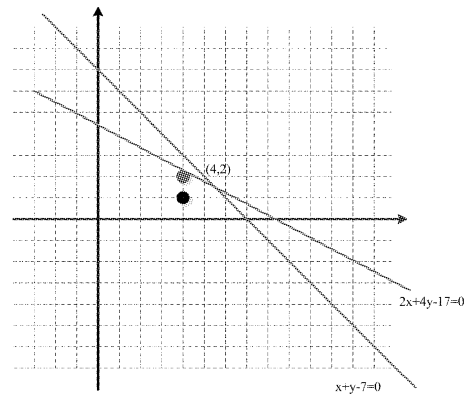


图 3 交互约束中原则 2 的应用(系统选择向用户展示的点)

系统给出交互元素(4,1), (3,2), (6,3)后, 用户选择其中最接近目标元素的元素(4,1), 用户目标元素的范围被进一步确定, 如图 4(a)所示。之后, 依据原则 2 和原则 1, 系统和用户之间继续交互, 最终找到目标元素(4,2)。在以上的示例中, 系统与用户交互 3 次后找到目标元素(4,2)。



(a) 用户在{(4,1), (3,2), (6,3)}中选择(4,1)后圈定的元素范围



(b) 用户在{(4,1), (5,3), (6,3)}中选择(4,1)后圈定的元素范围

图 4 系统和用户的交互过程

结束语 探索式搜索针对用户信息搜索需求模糊的应用而产生, 这种搜索方式比一般的信息检索更难以实现。交互是探索式搜索中必不可少的环节之一, 本文给出了探索式搜

(下转第 436 页)

## 参考文献

- [1] FELDMAN R, DAGAN I. Knowledge discovery in textual databases[C]//International Conference on Knowledge Discovery & Data Mining. 1995;112-117.
- [2] 刘静. 面向中文微博关键词提取技术研究[D]. 长沙:中南大学, 2014.
- [3] TAN P N, STEINBACH M, KUMAR V. Introduction to Data Mining[M]. Beijing: China Machine Press, 2010.
- [4] LUO S M, WANG Z K, WANG Z P. Big-Data Analytics: Challenges, Key Technologies and Prospects[J]. ZTE Communication, 2013(2): 11-17.
- [5] 陈晓云. 文本挖掘若干关键技术研究[D]. 上海: 复旦大学, 2005.
- [6] rickjin. 通俗理解 LDA 主题模型[EB/OL]. [http://blog.csdn.net/v\\_july\\_v/article/details/41209515?utm\\_source=tuicool&utm\\_medium=referral](http://blog.csdn.net/v_july_v/article/details/41209515?utm_source=tuicool&utm_medium=referral). 2014.
- [7] DAMILEVSKY M, WANG C, DESAI N, et al. Automatic Construction and Ranking of Topical Keyphrases on Collections of Short Documents[C]//SDM. 2014.
- [8] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003(3): 993-1022.
- [9] HOFMANN T. Probabilistic Latent Semantic Indexing[C]//ACM Proceeding of the 1999 ACM SIGMOD International Conference on Management of Data. New York: ACM, 1999; 50-57.
- [10] RAMAGE D, HALL D, NALLAPATI R, et al. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora[C]//Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing; Volume 1. Association for Computational Linguistics, 2009; 248-256.
- [11] LIU J L, SHANG J B, WANG C, et al. Mining Quality Phrases from Massive Text Corpora[C]//ACM Proceeding of The 2015 ACM SIGMOD International Conference on Management of Data. New York: ACM, 2015; 1729-1744.
- [12] ZHAO W X, JIANG J, YANG J H, et al. Topical Keyphrase Extraction from Twitter[C]//Proceedings of the 49<sup>th</sup> Annual Meeting of the Association for Computational Linguistics. 2011; 379-388.
- [13] HAN J, PEI J, YIN Y. Mining Frequent Patterns without Candidate Generation[C]//Proceedings of the 2000 ACM-SIGMOD International Conference on Management of Data. New York: ACM, 2000; 1-12.
- [14] LIU Z, HUANG W, ZHENG Y, et al. Automatic keyphrases extraction via topic decomposition[C]//EMNLP. 2010.
- [15] 郝峰. 文本关联分析中频繁项集挖掘的研究与改进[D]. 太原: 太原理工大学, 2008.
- [16] 蔡鹏飞. 基于概率图模型的关联规则更新方法与实现[D]. 昆明: 云南大学, 2013.
- [17] 李艳美. 基于贝叶斯网络的数据挖掘应用研究[D]. 西安: 西安电子科技大学, 2008.
- [18] 徐文海, 温有奎. 一种基于 TF-IDF 方法的中文关键词抽取算法[J]. 情报理论与实践, 2008, 31(2): 298-302.
- [19] YAN X, GUO J, LIU S, et al. Learning topics in short texts by non-negative matrix factorization on term correlation matrix[C]//SDM. 2013.
- [20] RAJARAMAN A, ULLMAN J D. Mining of Massive Datasets[M]. Cambridge: Cambridge University Press, 2012.
- [21] 章志刚, 吉根林. 一种基于 FP-树和数组技术的频繁模式挖掘算法[J]. 计算机工程与应用, 2014, 50(2): 103-106.

(上接第 402 页)

索系统中交互设计的基本原则, 并对其进行示例验证。交互有效性判定是一个难题, 本文给出了一种交互度量, 我们将进一步充实有效交互的理论成果, 给出交互有效性的定量的判定度量方式; 在本文的基础上, 可以结合已有对信息语义关联网的研究成果, 提出新的探索式搜索交互方法, 将本文思路应用到探索式搜索中, 检验其实践效果。

## 参考文献

- [1] 任磊, 魏永长, 杜一, 等. 面向信息可视化的语义 Focus+Context 人机交互技术[J]. 计算机学报, 2015, 38(12): 2488-2498.
- [2] FURNAS G W. A fisheye follow-up: Further reflections on Focus+Context[C]//Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems. Montreal, Canada, 2006; 999-1008.
- [3] ZHANG Y, LI Y, WANG H. A Study on Exploratory Medical Search Behavior Based on Cognitive and Log Analysis[J]. Library & Information Service, 2014, 58(11): 36-42.
- [4] TAN B, LV Y, ZHAI C X. Mining long-lasting exploratory user interests from search history[C]//ACM International Conference on Information and Knowledge Management. 2012; 1477-1481.
- [5] SUN H, JIANG C, DING Z, et al. Topic-Oriented Exploratory Search Based on an Indexing Network[J]. IEEE Transactions on Systems Man & Cybernetics Systems, 2015, 46(2): 1.
- [6] ZHAO Q, WANG C, JIANG C. Hsim: A Novel Method on Similarity Computation by Hybrid Measure[C]//International Conference on Information and Communication Systems. IEEE, 2015; 160-165.
- [7] MAO Y, SHEN H, SUN C. A Social-Knowledge -Directed Query Suggestion Approach for Exploratory Search[C]//International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery(Cyberc 2011). Beijing, China, 2011; 1-8.
- [8] KOTOV, ALEXANDER, BENNETT, et al. Modeling and analysis of cross-session search tasks[C]//Paper Presented at the Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. Beijing, China, 2011.
- [9] QVARFORDT P, GOLOVCHINSKY G, DUNNIGAN T, et al. Looking ahead: query preview in exploratory search[C]//International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2013; 243-252.
- [10] QU Y. Supporting Ideation by Integrating Exploratory Search, Browsing, and Curation[C]//ACM on Conference on Human Information Interaction and Retrieval. ACM, 2016; 361-363.