DOI:10.13530/j.cnki.jlis.184007

基于时态语义的 Web 信息检索实践进展与研究综述*

沈　思　李成名　吴　鹏

摘　要　基于时态语义的 Web 信息检索在动态时间信息挖掘、群体记忆、时间问答系统等检索情景中具有相对广泛的应用。在大数据和人工智能迅猛发展的大趋势下,对基于时态语义的 Web 信息检索从关键技术的角度进行系统而全面的综述,不仅有利于了解该领域研究的整体状况,而且有益于把握时态检索的未来发展趋势。本文在引入文本时间信息的抽取和标注并融合时间信息的检索模型的基础上,以时态语义的技术为整体脉络,从三个方面综述研究情况:信息需求中隐含的时间意图分析,加入时间因素的检索模型构建,时间对提升检索结果的生成。以时态语义检索的本源问题和其在学术文献上的相关应用为切入点,提出时态语义检索未来的发展趋势:识别多源异构信息下的时间表达,构建能识别查询的时间预测模型,搭建能精准检索时间意图的检索平台和开发基于深度学习的隐含时间意图自动分类模型。参考文献 91。

关键词　信息检索　时态语义　检索模型　时间戳　时间排序模型

分类号　G254.9

Review of Web Information Retrieval Based on Temporal Semantics

SHEN Si, LI Chengming & WU Peng

ABSTRACT

Web information retrieval based on temporal semantics is widely applied in the retrieval scenarios of dynamic temporal information mining, collective memory and temporal question answering system. By fully utilizing the temporal information contained in the Web documents, we cannot only retrieve the set of texts related to specified lexical semantics at a certain time point but also identify the pattern of theme evolution of contents in each independent information carrier over time. A total of 94 studies related to this topic are retrieved and analyzed. The Chinese search words are "时态语义", "时间信息检索", "时间信息抽取", "检索模型", and "深度学习"; the English search words are "temporal semantic", "temporal information retrieval", "temporal information extraction", "retrieval model" and "deep learning". The databases searched are Wanfang Database, CNKI, PQDD-ProQuest Digital Dissertation, ISI Web of Knowledge, Elsevier ScienceDirect, Academic Source Premier-EBSCO, LISA, SpringerLink and EI-Engineering Village

* 本文系国家自然科学基金项目"基于时间感知模型的学术主题检索与演化挖掘研究"(编号:71503124)的研究成果之一。(This article is an outcome of the project "Research of Temporal Information Retrieval and Temporal Dynamic for Research Topic Based on Time-aware Language Model" (No. 71503124) supported by National Natural Science Foundation of China.)

通信作者:沈思,Email: sszcgfss@gmail.com, ORCID: 0000-0002-6990-410X (Correspondence should be addressed to SHEN Si, Email: sszcgfss@gmail.com, ORCID: 0000-0002-6990-410X)

2. The extraction and labeling of text's temporal information and the retrieval model incorporated with temporal information are used. Existing studies are classified by comparison and content analysis and are reviewed in terms of each step of information retrieval in the narrow sense. Related researches in three aspects, namely, implicit temporal intention analysis in information requirement, construction of retrieval model incorporated with temporal factors, and promotion of retrieval result generation by temporal information are comprehensively reviewed under the framework of temporal semantics technology. Firstly, basic researches on automatic extraction of text's temporal information and retrieval model incorporated with temporal information are reviewed. A special focus is placed on studies related to temporal perception, retrieval and cognition, temporal query intent algorithms, methods to determine the temporal dimension of literature, temporal-based text similarity computing, and temporal ranking model. We also summarize four major future research directions in the field of Web information retrieval based on temporal semantics: multi-dimension time expression extraction, acquisition of implicit temporal intent, similarity computing in deep-level temporal retrieval model, and temporal intent classification based on deep learning. The major limitation of the present review is that besides the overall analysis of theories on Web information retrieval based on temporal semantics, there is also a need for systematic review of relevant theories and concepts. A systemic and comprehensive review of existing studies about Web information retrieval based on temporal semantics from the perspective of key technology is not only beneficial for knowing the research progress and status quo of this field but also helpful to grasping the future development trend of temporal retrieval in the era of big data and artificial intelligence. Such review can help us develop ways to incorporate the technique, methodology and concept of deep learning into information retrieval based on temporal semantics. Originality of this paper is manifested in two aspects. Firstly, key technology involved in Web information retrieval based on temporal semantics is taken as the entry point. Then we perform a systematic and comprehensive review on studies related to information extraction, organization, mining and presentation based on temporal semantics. Secondly, the future development trend of information retrieval based on temporal semantics is pointed out after literature review and a brief analysis on the latest development in this field is carried out. 91 refs.

KEY WORDS

Information retrieval. Temporal semantic. Retrieval model. Timestamp. Temporal ranking model.

0 引言

面对海量、模糊、复杂关联且动态变化的信息,用户一般很难用精确的查询式对检索对象的时间范围和语义精确描述的检索情景进行检索。一个检索系统如果能够获取网页当中的时间信息,这对于用户来说是非常有助益的。大部分常用的检索系统均不能精准地从网页中抽取时间信息,而且构建一个高效点对点的时间系统尚需要一定的时间。比如,自 1997 年起分别发生过"1997 年卡斯帕罗夫在纽约与'深蓝'对弈""2006 年中国象棋人机大战""2011 年'深蓝'的同门师弟'沃森'在智力问答节目中挑战人类冠军"以及"2016 年'AlphaGo'围棋比赛战胜人类"这四项人机大战事件。用户在输入"人机大战"这个查询词的时候,当前的搜索引擎 Google 和 Yahoo 给出的更多的是 2016 年的检索结果。这一检索结果是按照检索的热度或关注度给出的,而不是按照时间序列呈现的。另外,

涉及文件的时间戳和文献内容真实时间差时，对于非结构化文献，比如网页，从中抽取真实的时间是一个非常复杂的任务，比如一篇于 2012 年发表的文本内容却谈论 1997 年人机大战的报道。而检索者本意是想查询 2012 年与人机大战相关的内容，但目前的检索系统只能给出与论文发表相关的时间而不能真正地提供论文当中研究内容所涉及的时间。

在信息检索领域，基于时态语义的 Web 信息检索是正在兴起的一个研究领域，从用户信息需求的角度来看，是文献语境处理的重要维度之一[1]。根据信息理论，时间性或者时间流通性是决定一篇文献信度的五个关键要素之一，其他四个为相关性、精准性、客观性和覆盖度。从时间性或者相似框架的角度来看，检索时返回与时间相关的文献会呈现出检索结果的整体时间概貌[2]，这对于检索理解、检索消歧、检索分类和结果多样化的呈现也有一定帮助。从信息需求角度来看，基于时态语义的 Web 信息检索针对的是特定检索场景——“弱信息检索需求”[3]。在上述人机大战检索场景下，如果用户缺乏对所搜寻的人机大战对应事件时间点的“宏知识”，则仅能得到大量与近期“AlphaGo”人机大战的相关报道，无法满足用户对隐含时间信息的检索需求。

根据 Alonso[4] 等提出的基于时态语义的 Web 信息检索的研究目标，本文从时态语义检索探究的技术视角出发，按照狭义信息检索各主要步骤对现有研究进行分类和综述。在介绍相应的研究基础的前提下，分别综述了信息需求隐含时间意图分析与识别，加入时间因素的检索模型构建，检索结果中时间信息的呈现等各步骤中主要采用的方法和技术，并指出基于时态语义的 Web 信息检索研究未来的技术发展方向。

1 时态语义检索的研究基础

时态语义检索研究包含两项关键的基础技术：文本时间信息自动抽取和基于时态语义信息的检索模型构建。前者旨在从查询和文档结构不同位置中识别出包含时间语义的词汇和短语，而后者则希望在计算查询式和文档之间相似度时能够有效地利用这些时间信息。因此，在介绍时态语义检索的主要技术方法之前，先对上述两个方面的相关研究进行简要梳理。

(1)文本时间信息的自动抽取技术

文本时间信息是指科学文献、博客、微博、邮件和新闻等文档内容中所包含的时间区间，主要通过时间表达呈现出来。作为具有丰富自然语言形式的时间表达，实际上是带有时间意义的序列标记。关于时间信息抽取的研究主要是与时间表达的抽取和时间间的关系抽取相关。按照 Schilder 和 Habel [5] 的研究，把文本时间信息分成三类：显性时间表达、隐性时间表达和相关时间表达。显性的时间表达是在一个文献中被直接提及的并且能直接指定到某一个时间点和时间段上的时间，比如公历中的天或者年，例如“2010 年 7 月 4 日”和“2011 年 1 月 1 日”。该类时间表达在形式上具有明显的标记，在所表示的时间单位上比较精准，基本上是界定在时间的年、月、周、日、时、分和秒等单位上。由于显性时间表达具有明显的外部特征，因此对于这一类时间表达的抽取主要是通过基于规则的方法完成的。在覆盖度广、针对性强的规则下，该方法对显性时间表达的整体识别性能非常强。隐性的时间表达是指一个文档中不精确指定的时间点或者时间段，例如“2010 年的独立日”和“2011 年的新年”就是隐性的时间表达，该表达可以间接地转换为“2010 年 7 月 4 日”和“2011 年 1 月 1 日”。隐性时间表达在形式上不仅包含明显的时间标识，而且含有带有时间语义的节日、纪念日和特殊词汇，在所表达的时间上具有一定的隐含性，必须通过特定领域知识才能确定这一表达的精准时间。对于隐性时间表达的识别主要是基于统计和规则相结合的策略。规则的方法主要用于识别时间表达中具有明显标记的部分，而统计的方法则主要用于识别隐性的时间表达的部分。这种策略在

识别隐性时间表达的性能上与识别显性时间表达的规则方法相比有一定的差距。相关时间表达是指借用时间参考信息可以推断出某一时间点和时间段的时间表达,这一参考时间可能是在文档中或者是发表文献本身被提及的显性的时间,也可能是隐性的时间,例如“这个星期天”或者“下个月”就是一个相关时间表达。相关时间表达在形式上主要由年、月、周、日、时、分和秒等时间单位来表示,并且这些时间单位前面有一定的修饰词,如“这”“下”“上”“那”“这个”“那个”“下个”“上个”“这一”“那一”等,在所表达的时间上不仅具有隐含性而且需要上下文推理才能获得确切的时间。对于相关时间表达的抽取主要是基于机器学习的方法,因为这一类时间表达不仅没有显性时间表达的突出特征,也没有隐性时间表达的局部时间特征,所以只能根据相关时间表达中的分布词汇制定特征模板,通过隐马尔科夫模型、最大熵模型和条件随机场模型、神经网络模型等机器学习的方法完成相关时间表达的抽取。

时间间的关系主要是指不同时间表达之间构成的关系,这种关系主要由先后、包含和并存等类型构成。时间间关系的获取是构建在时间表达的抽取基础上的,因此时间表达抽取的精准度从根本上决定了时间间关系获取的准确度。时间间关系判定的方法主要由基于规则的方法和基于机器学习的方法两种策略构成。基于规则的方法主要是针对显性时间表达分布较多的文档,因为根据显性时间表达的特征,可以通过基于制定的时间转换规则,对文档中的时间统一处理成某一单位下的时间表达形式,然后通过比较时间的先后、包含与否和是否同一时间,最终实现对时间间关系的判定。基于机器学习的方法主要偏重于隐性时间或相关时间表达比较多的文档。首先,通过人工对文档中的时间表达进行先后、包含与并存等类型的标注;其次,选取相应的特征,执行特征验证并进而确定最优的特征;最后,通过所选取的机器学习模型,构建时间关系判定模型,从而自动对文档中的时间间的关系进行判定。

在时态语义检索中,文本时间信息的标注和抽取是重要的一环。具体由三个主要任务构成:第一是抽取或者识别时间表达;第二是归一化,对时间表达的不同形式统一化处理;第三是时间标注,对时间表达完成格式化标注。在文本集合中标注的结果通常按照 TimeML 格式呈现,主要为时间格式的标准化 XML 语言。时态语义检索需要对抽取和标引后的时间词展开语义探讨。一方面,对不同类别的时间词在检索中的重要性加以区分。另一方面,针对不同领域文本中时间词的特点,讨论决定文档或查询的不同时间维度的方法。

时态语义检索对时间词的分类,来源于时态数据库中所涉及的与时间有关联的相应问题。时间主要被区分为有效和执行两种类型。有效时间是指发生在真实生活当中的一段时间,而执行时间是指具体存储在数据库当中的事实时间。上述被区分的时间概念在基于时态语义的 Web 信息检索中被拓展成如下两个方面:文本语境当中时间的对应有效时间,这一时间是在网页当中被提及或者隐含所指的时间;执行时间,是与文档时间戳对应的时间,主要是指网页被创建、修改或者发布的时间。自然地,由于一个网页会涉及不同的时间点,因此其所关注的时间是一个时间段而不是单一的时间点。这个文档时间戳的时间区间被标注为[开始时间点,结束时间点],也可以标注成[ct, lmd]或[pt, lmd];对于文献的阅读时间和文献存在的年限这些额外的时间也作为本文考虑的时间类型。在网页搜索场景当中,一个文档的阅读时间与一个检索的查询时间是一致的,因为用户完成一个检索之后会即刻看到查询的结果。另外,一篇文献的年限是阅读文献与文献时间戳之间的差值。

另一方面,领域敏感时间标记指当时间标记被应用于不同的文本文档时,时间表达的抽取和归一化处理在质量上的差异。对于时间标注的整个任务来说,先决条件是与文本文件共

现的时间表达的抽取、分类和归一化处理，这三个子任务形成了时间标注研究的基本任务。比较有代表性的研究有：当使用时间标记 DANTE 去处理 Wikipedia 的文本而不是新闻文本时，Mazur 和 Dale [6] 发现在新的文本领域上整体性能有一定的削弱；Strötgen 和 Gertz[7] 对不同时间表达所在的文献进行了研究，主要的文献源包括 Wikipedia 文件、短文本和科学摘要，这三种文献在显性时间表达、隐性时间表达和相关时间表达的识别效果方面呈现出了不同的性能；为了更加充分地描述每一个领域和语言对的特性，Dias[8] 构建了一个基于时间的“本体—时间—感知”WordNet，在这个 WordNet 当中每一个时间同义词，比如“current”“commemorate”“expected”会自动地从四个维度上完成时间标记：非时间的、过去、现在和将来。

随着各种非结构化文本数量的激增，在已有对时间信息各种界定的标准体系上，目前的研究主要对不同领域的非结构化文本当中的时间信息进行多个层面、维度和角度的抽取。在抽取的过程中呈现出如下特征：首先，基于通用文本上的时间信息抽取算法、特征和策略，结合领域化文本中时间信息的分布特点，通过改进相应的模型参数，实现对领域化文本中时间信息的抽取；其次，结合已有的对时间信息标注体系不同维度的界定，基于不同时间标注体系在时间信息自动抽取上的整体性能，对时间标注体系的通用性、领域适应性和整体覆盖度展开各个维度的判定，以期能确定一个标记相对简单、体系相对完备、适应性相对强的标注体系；最后，在时间信息自动识别的模型当中融入外在的领域化知识，进而提供时间信息识别的整体性能，这些外在的领域化知识不仅包含了词表中的词汇、知识库当中的知识元、本体当中的语义关系等静态的知识，而且涵盖了从海量语料库当中抽取的词汇互信息、术语构成概率、组块分布熵等动态的知识。

(2)融合时间信息的检索模型技术

检索模型是指一种可以提供查询式和文档之间相似度的度量方法，该方法的处理对象是查询式 q 和文档集合$\{D_1, D_2, \cdots, D_n\}$，处理过程主要计算每篇文档 $D_i(1 \leqslant i \leqslant n)$ 和某个查询式的相似度 Similarity(q,Di)。从数学理论的角度来区分，常用的检索模型主要包括向量空间模型、概率模型、基于排序的语言模型、神经网络模型等。这些模型基于一种共同的理念：文档和查询共有的词项越多，认为这篇文档和该查询越相关。时态语义检索模型用到的数学模型主要有三类。①Pagerank 系列：在时态语义检索的过程中，基于 Pagerank 链接之间的相互指向关系，充分利用网页文本信息当中的时间信息对所链接的网页文本进行深入的分析，并且以时间为主要信息要素对网页文本的重要性进行加权。基于融入了时间信息的 Pagerank 模型，在对含有时间信息的网页文本检索时，便会优先把与时间相关的检索结果呈现出来。②语言模型序列：使用文献时间戳或者网页内容中的聚焦时间来构建排序模型，融入显性、隐性时间和相关时间的语言模型，称为时间语言模型。语言模型的构建均是基于不同词项在文档中出现的数量构建的，而与时间词在文档中出现与否没有关系。时间语言模型就突破了这一限制，基于不同的时间词项在文献中出现的次数来估计语言模型。时间语言模型构建的初衷就是对文档中所出现的时间词建模，并把每个时间词都视作加权的对象。在检索的过程中，突出时间词的重要性，从而在基于时间语言模型所获得的排序结果中，把与时间相关的内容优先凸显出来。③TF-IDF 和 BM25 模型系列：该类模型的特点是通过使用简单的 TF-IDF 变种和 BM25 模型，针对文档中时间表达进行相应的加权，同时获取与时间表达相关联的关键词。由于 TF-IDF 的主要功能是抽取主题关键词，同时通过在该模型当中融入相应的时间特征，在一定程度上可以探测某一主题的研究趋势。BM25 更多的是与检索的相关性关联，在获取文档中的相关关键词的基础上，通过在该模型中融入时间的属性，在特定的时间跨度上发掘某一研

究主题的连续性。

传统检索模型仅将文本时间信息作为众多搜索结果排序因子中的一项,例如将文档时间戳融入统计语言模型中,引入文本的时间先验分布来调整查询与文档的语义相似度[9]。随着全文检索和自然语言处理技术的发展,融合时间信息的检索模型逐渐转为对文档内容中时间项与查询相关性的研究。该研究领域具体探究的特征之一是如何利用自然语言处理完成对文本当中实体的标注,并通过与相关的时间关联,把关联的实体—时间特征融入检索模型当中,从而提高时间检索模型的整体性能。

2 技术视角下的时态语义检索实践进展

关于时态语义检索研究进展,有按照时间信息检索与相关应用和根据时间脉络与时间点对相关的研究问题进行综述两大类。与以往研究不同,本文按照信息检索顺序步骤中各环节所用到的主要技术,对整个时态语义检索的研究进展展开综述。

2.1 信息需求中隐含的时间意图分析

有些事实只有在特定的时间阶段才是有效的,如最近出版的书籍、最新提出的理论、某位学者最新的研究动向、当前美国总统的姓名、正在举办的奥运会等,这些主题都隐含了从某一时间开始到某一时间结束的非明确表示检索时间限定的需求。

(1)时间感知查询认知层面

这类研究主要探讨查询场景中隐含的时间意图主要有哪些类别,即对查询式中隐含的时间意图进行分类。在海量数据的背景下,用户的搜索需求不断增加,其中有一类查询场景中搜索意图与时间紧密关联,例如在微博搜索中,并非文档时间越新越重要,查询相关文档分布的高峰时刻,与查询主题具有更好的语义相关性,而这类查询场景被定义为时间感知查询。从检索主题上考虑时间,研究者尽可能扩展检索语言进而增强时间的语义性。关于查询式中时间意图的定义,Kanhabua[10]界定了明确的时间意图查询和潜在的时间意图查询的内涵与外延。通过对 AOL 查询数据的统计,Nunes[11]发现在所有的查询当中 1.5% 的查询是明确的时间意图查询。因此,明确的时间意图查询从显性的角度给出了特定的时间阶段所含有的具体日期,而潜在的时间意图查询则为指向某一个时间段或者不含有显性时间的内容,即用户在头脑中对该查询词进行时间界定,但在查询式构造时没有给出明确的时间表达限定。上述两位研究者通过相关的数据调查,对查询场景中的时间意图进行明确的界定和划分。更进一步,根据 Jones 和 Diaz[12]的研究工作,潜在时间意图查询可分为三类:非时间性查询,这一类查询与感知时间无关;时间无歧义查询,这一类查询的特点是具有精准的时间阶段;时间歧义性查询,这类查询与周期性事例或非周期性的事件相关。该研究从时间语义细颗粒度的划分角度,对潜在时间意图的查询进行更加细致的分类,为基于支持向量机等机器学习模型对潜在时间意图查询自动分类奠定了基础。基于所搜集的网络语料,Campos[13]对隐含时间的查询进行整体上的探究,发现 75% 的是非时间的,而 25% 的是隐含时间特征的。这一研究从具体定量的角度对隐含时间的查询分布情况进行细致的统计。

在上述对隐含时间查询界定和分类的基础上,通过对含有时间的网页查询实验,Joho 和 Jatowt[14]发现对未来的查询是最困难的,而对于过去内容查询的可能性也较弱,效果相对最好的是对当前信息的查询。在探测季节查询中,Shokouhi[15]设计了一个时间序列的分析方法,从查询隐含时间信息分类的结果的角度,对时间感知性的判断准确度展开了全面的探究。基于社会检索场景中的用户隐含时间意图,Khodae[16]提出了一个时间感知分类框架,命名为“时间感知社会信号系统”,根据社会行为的时间,把用户社会的兴趣点分为如下五类:当前、

正在进行、季节性的、过去和随机的。在这一分类框架的基础上，相关的研究者可以从时间的角度对人的社会行为进行历时、多维度的分析。针对社会搜索中用户隐含的意图，通过探测时间模式和时间动态的演化，Bannur 和 Alonso[17] 系统分析和探究了 Facebook 中的签到数据，发现呈现出非常明显的季节性特征。这一研究是基于时间感知分类框架进行的非常有代表性的探究案例。

在界定时间感知查询的内涵和外延的基础上，本部分对时间意图查询的不同类型进行了多维度分析，并重点从不同角度总结隐含时间的研究情况。如何从不同文本当中精准而有效地把隐含的时间意图挖掘出来，这一方面与所设计的时间意图分类框架有直接的关系，另一方面也取决于文本挖掘技术的发展水平。从目前整体研究的进度来看，时间意图的分类框架体系不宜过细，并且应当根据不同的文献类型制定不同的分类体系。

(2)时间查询意图的算法层面

在时间查询意图分类任务的研究中，研究者按照查询式的隐性或显性意图设计算法，并用该类算法把查询自动分到预先设定的过去、当下、将来和非时间等四个类别当中。通过贝叶斯分析，Gupta[18] 识别出了多维度的特征，并用该特征对时间敏感的查询进行了分类，这些特征是在 Jones 和 Diaz[19] 对时间特征的界定和探究基础上完成的。这一探究为如何挖掘时间的多维特征奠定了基础，但所使用的算法相对简单，整体精准性不高。Zhao[20] 基于 Wikipedia 这一特征信息源，探究了时间序列数据的使用问题，并按照时间查询意图进行分类探究。上述两个研究从相对宏观的角度探究了查询意图的分类问题。从整个流程的角度看，设计隐含查询意图的自动识别算法主要涉及如下三个方面的问题。

首先，检测一个查询是否需要按照时间敏感性分类展开特殊处理。从检索结果来看，Efron 和 Golovchinsky[21] 的研究表明，对隐含时间意图的查询式提高时间在近期或者特定时间段的文档的排序结果，会降低检索结果的准确率。从查询日志的角度，相应的研究者对查询意图完成了判断。Kanhabua[22] 基于查询日志的分析和相应算法可以判断某一个查询是否是隐含时间信息的查询。虽然这一探究具有可行性，但所判断的准确性有待进一步提高。在查询相应属性的基础上，比如周期、趋势和高峰等，Kulkarni[23] 通过分析大规模的查询日志，给出了一个查询意图是否为敏感性时间的程度。这一研究开创了使用日志大数据的研究先例，并且为判定查询意图是否为敏感性时间提供了可操作性的规范和流程。在搜狗实验室的查询日志基础上，张晓娟[24] 根据检索词和时间词共同出现的频次，提出了自动识别检索式的潜在时间意图的算法。Kanhabua[25] 对潜在的事件完成了分类，在分类过程中，根据信号处理的过程，对事件的时间序列查询特征进行发掘，并发现与事件相关的查询具有明显的时间意图。这一研究以事件为切入点，对查询的时间意图在新的视角下展开探究。基于从 Wikipedia 中获取到的时间数据，Zhao[26] 对查询完成了相应的分类操作，一方面通过 Wikipedia 中用户浏览的时间来消除 Wikipedia 百科词条的歧义性，另一方面基于特定的时间，对 Wikipedia 查询的意图执行消歧。该研究充分利用了所选择数据的不同时间维度，从消歧这一具体研究任务出发，对查询展开有针对性的分类操作，从研究方法上看具有一定的创新性。针对网络中的链接数据，Rahoman[27] 通过挖掘和利用隐含的时间特征构建了一个检索系统，并在检索系统中对相关的查询进行简单的分类处理，在一定程度上提高了这一系统的检索效率。

其次，决定是否返回更多历时或最近的信息来提供相关的结果。根据 Cheng[28] 的研究结论，时间感知查询可以划分成两类，一类是近期感知查询，即用户希望得到与主题相关的最新的返回结果，一类是目标感知查询，查询结果源自一个时间段。对于这两类查询，比较有代表

性的研究有以下方面。对于查询的新闻意图，通过查询的动态和点击的频率，Diaz[29]给出了一个有效的识别方案，该方案有机结合了查询动态和点击频率，在算法上把查询的动态和多维度的知识融入查询意图识别的探究中，具有一定的创新性。相对于无监督学习策略，Konig等[30]基于有监督学习的方法探测新闻检索结果的意图。这一方法虽然整体性能较为突出，但囿于有监督学习对特定语料的依赖性，这一方法的迁移性和适应性整体上不强。

最后，面向不同的时间类型抽取不同的时间特征。探讨如何对查询式和文本中能体现出隐含查询意图的时间词进行识别，抽取对象主要为查询式本身、文档内容、文档元数据。所抽取的时间特征信息源主要来自如下两类：第一类，可以从不同语料、不同领域文本中得到文档时间戳信息，比如《纽约时报》的标注语料库，该语料库涵盖了20年的新闻报道；第二类，可以从不同语料、不同领域的文本中得到聚焦时间的信息。这一领域有代表性的研究具体如下。通过模型拟合的方法，Boer[31]推断出基本的时间信息，并在艺术、历史和战争等历时的语料上抽取时间信息。这一研究虽然取得了一定的效果，但模型拟合的策略在一定程度上会造成数据过拟合现象的出现，从而造成对某一类时间信息识别的非常突出，而对其他特定的时间信息识别整体性较差。在对时间表达的极性和范围领域界定的前提下，Jatowt 和 Yeung[32]挖掘了新闻领域的时间表达特征，并得出新闻类报道中的时间特征相对容易获取的结论。通过利用时间表达的极性特征，研究者所提出的挖掘算法对极性时间表达的识别效果非常突出，而对于普通的时间表达识别效果有待进一步提升。针对微博用户的推送信息过载问题，结合登录的不同时间信息，De[33]提出了基于深度学习模型的推文（Tweet）排序策略。这一研究充分发挥深度学习的优势，提升了排序的精准度。围绕着维基百科中包含人名、地名和机构名等实体的文本，Kanhabua[22]完成了针对维基词条中与时间演化相关的时间信息抽取任务，并分析所抽取出的时间信息的特征。该研究不仅抽取了时间信息，而且所抽取的是与人名、地名和机构名相关的历时时间信息，因此这一研究具有一定的创新性。

在隐含查询意图的整个算法层面，目前已有的研究主要集中在有监督学习的模型上，无论是隐含时间信息的判断，还是查询词与时间词关系的判定，亦或是不同类别事件与时间意图的自动对照的判别，均是基于有监督学习的模型展开的。但如何探究无监督模型在隐含查询意图上的挖掘，是目前这一研究方向应主要关注的焦点之一；所使用时间特征在一定程度上决定了隐含查询意图自动判定的精准度，因此围绕时间特征抽取的全面性和精准性成为该研究方向的一个重点，并且针对领域化语料进行时间特征的挖掘在当前网络发展移动化、社交化和智能化的趋势下变得更加紧迫和重要。

2.2 加入时间因素的检索模型构建技术

由于本文限定在狭义的信息检索范围内，因此对检索模型构建过程中的组织和存储不予讨论，仅关注信息检索过程中的模型构建研究。

（1）决定文献时间维度的方法

在查询式构建之后，从文档的角度，为了计算查询和文档的时间相似性，首先需要确定文档的时间信息，即确定文档的时间戳和文档聚焦的时间。检测决定文献时间维度的处理过程，这一处理过程对于改进时间检索的任务具有一定的促进作用，这一任务包括抓取、索引、查询理解和网页检索结果的排序。在文献的时间维度判定上，Kanhabua 和 Nørvåg[25]观测到，这一判定过程可被分为如下两类：决定文献的时间戳（即文献被创建、发表或者最后更改的时间）和决定文献的聚焦时间（即与内容相关的时间）。这一探究从整体上为文献时间维度的判定勾勒出了相对清晰的框架，在这一框架的基础上，后续的研究者对文献时间维度的判定展开了相应的探究。首先，Toyoda 和 Kitsuregawa[34]完成

了对文献时间戳的测定，基于链接网页的数值，提出了一个用来判断文献创建时间的测度方法。无论是文献时间维度框架的勾勒还是文献时间戳的判定均是通过使用临近网页来估算文档创建的日期或者最新更改的日期。这些探究均是建立在从网页当中可以抽取出时间信息这一假设前提下的，因此存在如下不足：第一，需要有一个大的外部文献集合以备计算来用；第二，在外部资源集合当中更改时间不能得到有效的保障，在一定程度上会削弱计算的精准性。针对上述问题，基于时间语言模型，在新闻报道的语料上，Kanhabua 和 Nørvåg[35] 通过检测非时间戳中的时间来处理这一问题。这一探究有效地解决了语料规模过小的问题。Jatowt[36] 通过多个渠道的数据源确定文档的时间戳，主要包括 HTTP 当中最新修改的信息资源的内容、被 Google 刚刚抓取过的最新时间、在任何公共网页文档中第一次出现的时间、第一次被 Twitter 标注的网页时间等。这些异构的数据源不仅解决了数据规模偏小的问题，而且在一定程度上可以解决数据时间更改的问题。基于无监督和有监督的算法，Garcia-Fernandez[37] 开发了一个自动判定法国历史文献发表日期的系统。这一研究虽然融合了无监督和有监督的算法，但在如何发挥两种模型的整体性能上确有较大提升的空间。为最大限度地减少时间因素的影响，Fukumoto[38] 通过文本自动分类的策略来判定文档中的时间戳。这一研究巧妙地把对时间戳的判断任务转化为了分类的问题，在方法上具有较强的创新性。针对非常规的文献发表时期，Peetz[39] 定义了一个时间阶段的突发点，并从查询的角度提出了一种算法用来确定初始检索文献集合当中的时间突发点，以便更好地判定文献集合的时间戳。从查询的角度，结合时间段的突发点来判定时间戳，在一定程度上可以解决文献时间更改的问题。在 Tweet 语料上，Ganesh[40] 通过融入历时的时间信息设计了一个新的表示学习模型，该模型整体上实现了时间与用户生成内容的有机对应。这种方法充分利用用户生成内容这一锚点，确定时间更改的具体时间线，进而提升时间戳判定的准确性。

较多的方法被用来判定文献的时间戳，鲜有方法用来判定文献聚焦的时间，特别是在缺失或者缺少文献内容时间表达的前提下。Jatowt 和 Yeung[41] 提出了两种估测文献聚焦时间的方法。这两种方法均是基于规则完成的，虽然在判定文献聚焦的时间上相对精准，但其规则的覆盖度整体上较差。2011 年，Jatowt 和 Yeung 已提出了页码聚焦时间是时间周期集合的这一理念[42]，并基于从大规模新闻文本中获取的数据，通过统计的方法对文献聚焦时间进行了判定[43]。相对于基于规则的方法，统计的方法在整体性能上更为突出。在微博和社会媒体上，Jatowt 和 Antoine[44] 提出了记忆和期望感知这一概念作为社会感知的一种补充，并把这一概念融入微博和社会媒体的聚焦时间判定上。虽然判定的整体性能不是太高，但把认知的相关理念引入到聚焦时间的判定上具有一定的创新性。通过设计一个相对精准的算法，Zhao[26] 把新闻报道的显性时间表达和隐性时间表达集中在一个统一的表示集合中，并对时间表达进行归一化处理，进而实现了对新闻报道中聚焦时间的判定。该研究充分利用了显性和隐性时间表达的特征来提升聚焦时间判定的精准度，但整个算法在时间效能上相对较差。基于大规模语义标注的文献集合，Abujabal[45] 提出了通过计算时间与命名实体在不同句子当中的共现频次，进而来判定文献聚焦时间的策略。该策略充分利用时间与实体之间的语义关联知识，有效地提升了聚焦时间判定的召回率，但在精准性上则过渡依赖于实体识别的效果。

在判定文献的时间戳的研究中，基于当前的海量数据源与大数据技术，以及对实时时间的精准应用，一定程度上实现了对时间戳的判定。但对于文献聚焦时间来说，在文献当中由于可能存在多个主题，这一探究具有一定的挑战性。可能的一种解决方法是基于所讨论的不同主题把文本分成不同的部分，文本的每一部

分可以被分配到不同的时间阶段当中,即对长文本来说,文本可能表达多个主题,对应每个主题存在一个最相关的聚焦时间。目前的研究主要根据文档内容仅对单文档所聚焦的时间执行探测,对长文档中单文档存在多个主题、多个聚焦时间应如何识别和处理的问题尚待进一步探究。

(2)时间文本相似度技术

基于字面匹配算法的查询式与结果文本的相似性计算在一定程度上会受到查询式语义变迁的影响。受这一影响比较典型的领域一个是检索概念更名,另一个是用于检索的词汇发生概念变迁。通过引入一个隐马尔科夫模型,结合测量两个术语跨时间的语义相似度和比较所获取的共现测量语境,Berberich 等[46]相对完整和系统地解决了上述两个问题。虽然在语义的判定上使用了概率统计模型,但该模型的性能整体上不强,需要进一步提升和改进。Kanhabua 和 Nørvåg[35]针对检索概念更名这一问题,基于维基百科的内容,给出了一个可以自动创建同义词实体的方法。该研究充分利用了维基百科中有关概念的历时演化信息和知识,根据实体的不同演化过程,构建实体的同义词集,在使用维基百科资源上具有较强的创新性。根据命名实体与时间周期的关联性,Tahmasebi[47]通过命名实体在时间周期序列下的语义演化,来推断不同时间下的语义内涵。这一研究以时间为维度,通过使用关联规则来探究实体语义变迁,在研究方法上具有一定的创新性。通过使用关联规则挖掘的方法,Kaluarachchi[48]挖掘了不同时间周期上与所使用的语义度等同的概念。通过引入语义度的概念,更加细化和全面地测评了与时间周期相关词汇的语义颗粒度。为了探测目标时间当中的查询,Zhang 和 Jatow[49]探究两个术语之间的跨时间语义相似度计算问题。从查询的角度,该研究针对术语的语义分布,探究术语之间在时间序列上的相似度问题。为计算与时间相关的词汇的语义变迁情况,Berberich[50]不仅利用了词汇本身的语义知识,而且结合了词汇在语境中的分布,并给出了具体的度量流程和框架。对于如何充分利用词汇的语境信息提高判定词汇语义度的精准性,这一探究给出了非常有价值的尝试。对于学术文献中具体内容的时间与出版时间差距比较大的文献集合,Radinsky[51]提出了一种通过计算不同词汇的隐含时间语义知识,按照各文献中出现时间构成的时间序列相似性分布情况,确定具有时间语义关联的词汇对。相较于上一研究借助语境的计算,这一研究更加细化了时间语义关联词汇判断的语义颗粒度。在确定的某一个时间段内,Chundi[52]发现时间所对应的查询关键词,更多的是与文献发表的时间有关联,而与关键词所聚焦的时间没有太大的关联性。这一研究从算法层面对文献发表时间与研究内容聚焦时间之间的关系进行了多方面的探究,在一定程度上拓宽了研究的领域。

通过分析与时间信息检索相关的研究,发现在时间文本相似度计算上,占主导地位的研究趋势是深度结合自然语言处理中的相关技术,结合领域词汇资源和词汇的语义知识,从历时和动态的角度完成对查询式和结果文本的相似度计算。在某种程度上,把时间文本相似度的探究往深层次、多维语义的方向推进。

(3)时间意识的排序模型技术

按照是修改了检索模型本身还是仅在模型的特征中加入时间属性,可以对检索模型进行类别划分。不改模型,仅在排序模型参数中加入时间属性,如下文的类别 1 和类别 2。直接修改模型本身,如时间语言模型,按照是只用到了时态语义检索中的文献时间戳信息,还是同时用到了聚焦时间和文献时间戳两种信息,得到类别 3 和类别 4。

类别 1:使用经典排序模型,通过查询反馈等方式,在原有排序结果的基础上,突出某些可反映用户隐含时间信息的文档。在 Google 学术搜索当中,用户可以直接把检索限定到某一特定的时间段当中,返回的结果具有较强的覆盖度和精准度。基于 TREC 2011 的微博数据集,

在对前 n 个查询微博排序的基础上，通过设定关联的信息在特定时间周期内爆发的界限标准，结合所增加的时间衰退因子，Wei[53] 调整了与检索相关的关联数据值，调整后的检索效果在精准度上有了非常大的提升。针对文献中的时间点信息，Setty[54] 设计了一个名称为 INZEIT 的系统，该系统的主要功能是通过检测有关给定查询深度的时间点，用来辅助用户实现与时间点相关的结果的检索和排序。在融入时间点这一特征知识之后，虽然检索的召回率提升了，但整个检索的速度有一定的降低。采用 RankSVM 等相对成熟的机器学习排序方法，Dong[55] 把从 Twitter 当中抽取的新的 URLs 融入通用的网页检索系统当中，实现了与时间有关的结果的检索和排序。该研究的创新之处在于把机器学习的方法融入排序当中，确保检索效果的精准性。基于时间点击特征集合和查询重构序列集，Inagaki[56] 通过利用点击返回的特征来判断时间序列上文献的关联性是如何演化的，并对时间序列上的文献排序。这一研究充分利用了实时的时间信息，从时间排序的结果上看，提升了排序的精准度。同样，结合查询日志当中的时间点击信息，Chang[57] 对检索结果重新排序，与上述研究的不同之处在于时间单位限定在一天之内。在检索性能上，虽然精准性有所提升，但由于所限定的时间比较固定，整体检索效率有所降低。在把输入看成隐含的时间查询并依据从文献当中抽取出时间特征的基础上，针对新的文献所给定的排序提升需求，Zhang[58] 给出了一个重新排序值调整的策略。该排序策略有效地提升了排序的精准度，实现了对新文献的精准排序。利用简单统一的体系结构，Meng[59] 基于 LSTM 这一深度学习模型，提取了网页评论中的问答时间信息。这一研究通过 LSTM 寻找时间信息识别的特征，提升了识别的准确度。在将信息出现的时间因素作为性能指标对 Web 搜索得到的文档排序的基础上，王文[60] 提出了一种基于舆情的敏感新信息检索算法。该算法提升了检索效果的召回率，但在对新信息的排序上则有所保留。针对网页时态结果的排序，李筱文[61] 提出了 CT-Rank 和 NTLM 两种算法。这两种算法，前一种排序效果的精准性有所改进，但整体幅度不大；后一种算法，在融入时间之后，排序结果有了较大的提升。在排序效率上，两种模型相差不大。

类别 2：链接技术和方法在时间检索模型当中的应用。结合融入了时间因素的数据集合，通过系列的检索实验，Baeza-Yates[62] 展示了 PageRank 算法在检索过程中的时间偏差，并得出 PageRank 算法对于提升新的刚创立的网页是无效果的结论。这一探究不仅得出了有效的结论，而且为后续的探究奠定了基础。基于链接分析，Berberich[63] 介绍了两种方法，即 T-Rank Light 和 T-Rank，这两种方法不仅考虑了更新度和活跃度，而且融入了检索当前文献的链接，在这一研究当中更新度是指更新的时间邮戳，活跃度是指更新的频率。所提出的两种算法，前一种算法在检索的速度上要优于后一种算法，但在检索的精准性上后一种更加突出。基于网页链接结构特征，特别是其时间演化的具体分布情况，Cho 等[64] 提出了一个新的排序指标用来解决 Baeza-Yates 所提出的问题。与 Baeza-Yates 所得到的结果相比较，这一研究在一定程度上给出了新的解决方案。通过探测网页的变体和链入在时间上的更新度，Dai 和 Davison[65] 估测了网页的授权情况，并把这一时间更新度的信息融入到时间排序概率模型当中。与未融入时间更新度的情况相比较，这一概率模型所得到的结果在精准性上有较大幅度的提升。在使用文献当中的文本、时间、地理查询术语的临近值和链接值的基础上，Strötgen[66] 提出了一个新的排序算法。该算法充分利用文献中的实体特征知识，不仅有效解决了排序结果的精准性，而且由于预先筛掉了一部分不需要的排序信息，排序的整体效率也有较大的提升。在 RankSVM 等排序学习算法模型的基础上，结合链接的方法体系特征知识，Kanhabua[67] 提出了一个未来预测的排序模型。从所获得排序模型

的效果上看,相关链接特征知识的融入在一定程度上提升了排序模型的准确度。结合网页的链接信息,Campos[68]开发了一个语言独立模型,该模型主要用来处理时间维度上的查询,并可以把最相关的时间周期识别出来。从时间周期这一维度入手,该研究提出的语言独立模型不仅确保时间周期结果的准确性,而且整体效率与未处理时间维度上的查询模型相差不多。基于静态关联的文献,融入链接的策略,Gupta[69]提出了一种算法,该算法主要是使用时间表达来对检索结果进行多样化的排序处理。从排序性能上看,这一算法一定程度上确保排序结果的召回率,实现了对排序结果的多样化处理。

类别3:将用户隐含的时间信息融入检索模型本身之中,主要使用的是文档外部的时间信息。在所涉及的模型中主要涵盖改进的PageRank模型、语言模型、tf-idf和BM25模型等几种模型。通过设定一个非固定的阻尼因数,Li等[70]改进了PageRank,该阻尼因数是受一个函数控制的,而该函数主要是依据消失的时间完成计算的。通过分析改进后的模型,由于融入了事前的时间信息,与未改进的模型相比,在精准的性能上有了较大的提升。Efron和Golovchinsky[21]提出了一个扩展的查询相似模型,该模型不仅考虑了文献发表的时间,而且考虑到了发表时间与查询之间的关系。这一扩展的查询相似模型,由于融入了文献发表的时间及其与查询之间的关系特征,在整体的检索性能上,精准性与速度均有一定的提高。不过上述三种模型都将更新度作为权重的最高因子,认为文档创建时刻越新,文档的排序结果应该越靠前。与此相反,Dakka[71]对不同查询主题区分对待,在排序中将时间越靠近查询主题热门时刻的文档看作越重要的文档。以热门时刻为时间锚点,通过细分的主题查询,这一研究方法有效地提升了查询的精准性,并且通过对不同主题有区别的查询,相对加快了检索的速度。围绕博客这一数据源,Zahedi[72]通过使用查询中的时间属性词,对检索出来的博客中的结果进行排序。从时间排序的结果上判定,所得到的排序结果确实在精准性上有了较大的提升。通过分析微博检索中时间信息的重要性,卫冰洁[73-74]提出了一种以统计语言模型为基础,融合Hashtag、聚类和时间信息三因素的排序算法。从检索的性能上看,所提出的语言模型由于有效地融入了不同因素,与未融入这些因素的模型相比,在精准率、召回率和排序的速度上均有较高的提升。这一类研究主要是围绕着如何使用时间这一关联维度展开的,均只利用了文档创立的时间,对于深入利用文档中的时间均未涉猎。

类别4:改进了基本的检索模型,同时融合了文本创建时间和文本内容中的时间信息。对标准的搜索引擎来说,从文献分布的特征来看,不仅文献的创建时间或者文献最后改变的时间是关联的,而且文献的时间内容也被拆分了。因此,时间检索结果的评分通过文献当中的所有时间表达来裁定或者通过验证文献的聚焦时间来判定。首次把文本和时间集中在一起对搜索引擎描述的是TISE系统,该系统为针对一个面向汉语网页内容的时间搜索引擎,兼容特殊时间区间的文本查询和时间查询这两种类型。从检索性能上看,这一系统突出地确保了查询的召回率,实现了检索结果的全覆盖。在《人民日报》新闻报道数据集的基础上,Lin[75]做了相似的工作,并且构建了名称为TASE(时间感知的搜索引擎)的原型系统。与上述系统相比,这一系统更加强化了检索的召回率,同时也在一定程度上改进了精准率。Vicente-Diez[76]搭建了一个西班牙搜索引擎,其创新点在于该引擎在计算相似度之前分别对查询式和文档集的时间词归一化。这一归一化处理,不仅提升了搜索引擎在时间检索结果上的精准性,更加大幅度地改进了检索的速度。Arikan[77]的研究则对文档内容中出现的所有时间词一视同仁,同时参与排序。这一策略确保了排序结果在召回率上的提升,使排序结果更加全面和系统。通过寻找与时间相关的文献,Berberich[78]提出了一种检索模型,该检索模型的创新之处在于充分利

用文献中时间表达的不确定性这一特征。从检索性能的角度分析，通过这一方法所获得的时间检索结果，具有最大程度上的覆盖度，有效地确保了检索的召回率。在微博这一语境中，通过遵照查询独立性的方法，Efron[79] 提出了一种竞争的分析策略，该策略能把当下的时间信息融入文献排名当中。从时间检索的整体性能上分析，对于时间单位的某一刻，所获取的检索结果具有较强的正确率。通过结合文献和查询之间的时间范围，Brucato[80] 改进了传统的排序模型，并提出了非对称度量空间的理念，对于不同检索任务的时间关联性进行了模型化的统一计量，比如，历史书籍当中的某一章节可能会指向某一个过去的时间周期，而来自 Twitter 社会化网络中的微博则相反会有一个非常窄的时间范围。这一改进的模型，有机地从历时的角度，确保了时间序列上检索结果的精准率和召回率。通过在时间感知排序算法当中融入文本和时间相关特征的策略，结合大规模数据集，林盛[81] 开发了一个时间感知的检索原型系统。由于这一原型系统充分利用了时间的信息，在检索的整体性能上有机地确保了检索召回率这一性能的提升。

根据时间属性在排序模型当中使用的程度，本文对与排序模型相关的研究完成了梳理和归类。从所使用的技术难度和深度来看，随着排序模型当中时间属性融入的程度越深，所需要的技术和模型越强和越复杂。基于排序的评价指标，在深入分析四种类别的模型之后，本文发现这四种类别的排序模型有一个共同需要改进的地方，具体如下：当处理时间查询的时候，时间和主题的测度值通常会作为查询时间和文本的一部分被计算，特别是对于计算整个关联值的时候，因为查询的文本和时间的部分是相互独立的；但在语义上文本和时间部分的查询是很难彻底被分开的；同时，在排序算法设计当中，当所描述的能满足显性时间信息需求的方法随着所研究的问题变化时，相应的算法拥有同样的缺点，即文本和时间查询部分被认为是独立的。这一问题在后续的时空信息检索模型中会得到一定的解决。

2.3 提升检索结果生成的技术

在查询的过程中，需要返回相关的文献，该文献能够满足时间的查询意图。这是一种新的呈现检索结果的方式，在有关时间特征问题的基础上，这一检索结果能够满足用户的信息需求。这类检索结果的时间主要包括了如下内容：网页片段的时间、网页结果的时间聚类、时间的多样性和时间聚焦结果的可视化等。

首先，网页片段的时间方面。目前的研究主要与构建高性能的针对检索结果的查询聚类相关。代表性的研究有如下方面。Alonso[82] 介绍了时间网页片段的概念，具体为文本部分被一定量的关联时间表达取代。该研究更加全面和系统地介绍了时间网页片段这一概念的内涵与外延，并从关联时间表达的角度探究如何挖掘时间网页片段中所包含的时间知识。另外，Svore 等[83] 分析了网页短文本中所包含的当下时间语境，这一研究的结果表明，从用户的角度分析，如何从网页短文本中挖掘相应的时间知识对于趋势的查询、展示新的时间内容是非常有帮助的。这一探究从方法论的角度给出了短文本中时间信息挖掘的方法和策略。

其次，网页结果的时间聚类方面。这一研究主要有两类。第一类研究考虑了与能找到的日期相关的所有时间表达，并且用相应的算法聚类。相关的研究工作由 Alonso 和 Gertz[84] 首先发起，这一研究用时间属性作为向量来表示一篇文献，该属性是从元数据当中抽取出来的。在充分利用时间属性特征知识的基础上，所获取的聚类结果得到了极大的提升，也充分体现了聚类中时间的属性。Alonso[85] 提出了 TCluster 聚类算法，是一种重叠的聚类算法，在这种算法当中每一个文献都是与时间文献相关联的，并且该时间文献含有一个三元组的列表，即<E,C,P>。这一研究一方面利用了时间文献三元组的特征知识，另一方面给出了融合时间表达特征的聚类算法，确保了整个聚类结果的

高性能和稳定性。第二类研究是通过检索得到大量的相关时间,而不是先前所考虑的与日期相关的时间表达。代表性的研究是由Campos[86]提出的,在网页片段当中识别相关的时间表达,在具体实现的过程中,如果文献共同使用了某一年的时间信息,这些文献就被聚类到一处。但是,由于该方法所得到的类别是基于通常的日期完成的,可能在主题相关上相对弱一些。但通过对时间表达的识别并使用以年为单位的时间特征聚类,在方法上具有一定的创新性。

再次,时间的多样性方面。如果一个查询具有多维度的解释,则该查询是歧义性的,只有一个会被认为是相关的。为了满足使用的基本需求,内容检索结果的分类在最大程度上对检索内容完成了有效的划分。从用户体验的角度分析,Berberich 和 Bedathur[87]认为查询的单一时间兴趣点结束了对返回文献的主导,检索结果的时间多样性是一种有效表达。这一研究从用户体验的角度对检索的多样性完成了实证性的探究,为后续的相关探究奠定了坚实的基础。与 Berberich 和 Bedathur 的工作形成对比,尽管歧义查询的时间多样性需要检索结果,Whiting[88]提出了在尽可能多的检索内容前提下的交错结果覆盖度概念。这一概念对评判检索结果的多样性具有较强的可操作性,为时间角度下的多样性检索结果的优化提供了切实可行的评估方案。

最后,时间聚焦结果的可视化方面。在通过使用时间信息来探究检索的研究上,对于展示文本当中的时间信息,很多可视化的系统已经出现。围绕着时间数据可视化的最早研究是由 Aigner[89]提出的,其研究开发了相应的模型并提出了基本的原则。该研究为后续时间维度下的可视化探究制定了切实可行的原则。Odijk 等[90]提出通过把信息检索和可视化界面融合的方式来支持纵向文献集合的探究。这一研究有效地解决了对纵向文献检索结果可视化探究的难题,具有方法上的创新性。在上述研究的基础上,Tran[91]将该问题与文本时间特征的利用联系起来,提出了面向历史事件的时间上下文语境化重构的方法,以提高对事件时间表中重要历史新闻事件的检索准确性和可视化呈现的多样性。该研究充分利用了文本中的时间特征,并有机地把这一特征与上下文语境进行了密切的融合。

以时间为主要维度呈现检索结果,不仅能够历时地刻画出检索结果的多样性,而且能够细颗粒度地呈现不同时间结点上的内容。但从目前已有的研究成果来看,由于文献的时间结果呈现一直处于探究的阶段,对于文献当中的时间信息哪一种是最好的呈现方式,文献列表、时间轴线、时间或者主题聚类,或者通过使用编码的时间信息术语云,这些问题一直没有得到有效的解决。

3 未来发展的趋势

在总结已有研究的基础上,本文提出时态语义检索未来的发展趋势主要集中在如下几个方面。

首先,如何将文本内容时间信息和学术文献检索相结合,从文献所有的时间表达中识别出相关的时间表达。尽管大部分研究工作能够找到与事件相关的兴趣时间点,但在探测与事件相关的时间时有一定的不确定性。由于很多时间表达是与一个事件相关联的但不是均等相关联的,所以判定相关的时间表达是一项有挑战的研究任务。因此使用基于文献、基于句子和基于特定语料库等三个不同语境维度上的特征知识成为一种趋势。比如,在学术文献中,时间表达“最终”在字面上是一个表示时间涵义的词汇,但该词及其被修饰对象在语义上是否表征时间信息,需要结合语义进一步分析与判断。如对于学术文献《突显人性化管理,构建和谐图书馆》的摘要片段“主要阐述了和谐图书馆的内涵……从……建设,形成……方面论述了……最终达到实现图书馆的和谐发展”中出现的隐性时间表达中的时间词“最终”而言,该时间词

描述了构建和谐图书馆的终极目标，因此能为时态语义检索提供有价值的时间信息。与此相反，出现在文献《国外网上文献传递服务系统的发展现状及特点》的摘要文字“……直接面向最终用户，努力减少中间环节……”中的“最终”一词，和其修饰对象一起构成在语义上与时间毫无关联的概念“最终用户”，在用于时态语义检索模型前需要予以剔除。

其次，在查询式隐含时间意图识别方面，基于文献所共有的时间特征，两个查询能认为是相似的吗？这一问题可以通过“战争”和“和平”这两个查询来予以说明，因为这两个词汇在特定的时间内是相关的，尽管它们通常出现在不同的文献当中。在这一研究领域，一个可能的应用是查询扩展。对于时间周期查询扩展来说，开发用来识别查询的时间预测模型是非常重要的，这方面目前还没有相应的探究，主要是因为目前的系统继续采用单一的预测方法，从而导致主要使用了时间表达当中的时间点而不是时间周期。在这一背景下，从时间周期当中获取相应的实体成为一项非常有意思的挑战，在未来几年的研究中，这一领域会受到更大的关注。例如，查询词“奥巴马”可能的一个周期查询的集合如下：“奥巴马，1961—2003”“奥巴马 伊利诺伊 参议员 1997—2004”或者“奥巴马 总统，2008—2012”。

再次，在检索模型的语义相似度计算中，检索结果中所包含的时间表达并不总是反映时间意图。例如，当检索关于某一个特定主体的未来信息（比如“安培经济学”）并且构建了一个查询式（安培经济学最新），用户常会得到一个含有词汇“最新”的网页，却是多年前推出的一个网页。因此，需要确保一个检索系统能够精准地把用户的时间意图翻译成相应的查询，并能检索出相应的文献，且该文献含有某一时间区间内的信息，而不是获取到只含有时间表达式术语内容的文献。

最后，针对非结构化文本，基于深度学习对隐含时间意图进行更加深入和精准的自动分类研究。总体上，对于不同用户检索行为的时间模式，与事件相关的信息需求在查询流中是可以被观察到的，比如流行事件的最高峰、重复事件的周期峰值。但是用户检索不流行的事件也是比较平常的事情，在查询流上这些事件并没有展示出时间的变化情况，比如刚发生的过去事件，由周年纪念日或者相似的事件被激活的历史事件，将来预计会发生的事件。为了解决动态类型时间探测的挑战问题，在未来的研究中可以引入深度学习模型，该模型能相对高效和精准地把给定的查询自动分到预先设定的多类型时间集合当中。

4 小结

随着图书、历史资料、数据档案等的数字化和电子化，对文本时间特征的挖掘和应用已从单纯的新闻报道，逐步转向面向其他多元的信息载体中的文本时间信息的组织与利用，而第三代搜索引擎提出了时态语义检索模型，该模型通过更充分地利用 Web 文档中的时间信息，结合人工智能的技术和方法，实现在指定时间点搜索指定词汇语义相关文本集的检索需求。在这一大的时代发展背景下，以时态语义 Web 信息检索中所涉及的关键技术为主线，对相关实践进展和研究进行了综述。从准确抽取学术文献当中的时间表达，构建用来识别查询的时间预测模型，精准计算检索模型的语义相似度和基于深度学习模型的隐含时间意图的自动分类等方面，对时态语义 Web 信息检索未来的研究趋势进行了展望。

参考文献

[1] Manica E, Dorneles C F, Renata Galante R. Handling temporal information in Web search engines[J]. ACM

SIGMOD Record, 2012, 41(3): 15-23.

[2] Kanhabua N, Blanco R. Temporal information retrieval[C]// 39th International ACM SIGIR Conference on Research and Development in Information Retrieval .Geneva, Switzerland , 2016:1235-1237.

[3] Palmer C L. Research practice and research libraries: working toward high-impact information services[EB/OL]. [2017-08- 09].http://www.ideals.illinois.edu/handle/2142/9742.

[4] Alonso O, Strötgen J, Baeza-Yates R A, et al. Temporal information retrieval: challenges and opportunities[J]. TWAW, 2011(11): 1-8.

[5] Schilder F, Habel C. From temporal expressions to temporal information: semantic tagging of news messages [C]//Proceedings of the Workshop on Temporal and Spatial Information Processing. Stroudsburg, PA, USA, 2001: 9-16.

[6] Mazur P, Dale R. WikiWars: a new corpus for research on temporal expressions[C]//Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Cambridge, Massachusetts, 2010: 913-922.

[7] Strötgen J. Domain-sensitive temporal tagging for event-centric information retrieval[EB/OL].[2017-05- 09]. http://www.ub.uni-heidelberg.de/archiv/18357.

[8] Dias G H, Hasanuzzaman M, Ferrari S, et al. Tempowordnet for sentence time tagging[C]//Proceedings of the 23rd International Conference on World Wide Web.New York, USA, 2014: 833-838.

[9] Li X, Croft W B. Time-based language models[C]//Proceedings of the Twelfth International Conference on Information and Knowledge Management. New Orleans, USA, 2003: 469-475.

[10] Kanhabua N, Nørvåg K. Learning to rank search results for time-sensitive queries[C]//Proceedings of the 21st ACM International Conference on Information and Knowledge Management. Hawaii, USA, 2012: 2463-2466.

[11] Nunes S, Ribeiro C, David G. Use of temporal expressions in web search[C]//European Conference on Information Retrieval. Springer, Berlin, 2008: 580-584.

[12] Jones R, Diaz F. Temporal profiles of queries[J]. ACM Transactions on Information Systems (TOIS), 2007, 25(3): 14-19.

[13] Campos R, Jorge A, Dias G. Using Web snippets and query-logs to measure implicit temporal intents in queries [C]//SIGIR 2011 Workshop on Query Representation and Understanding. Geneva, Switzerland ,2011:23-29.

[14] Joho H, Jatowt A, Roi B. A survey of temporal web search experience[C]//Proceedings of the 22nd International Conference on World Wide Web. Rio de Janeiro, Brazil, 2013: 1101-1108.

[15] Shokouhi M. Detecting seasonal queries by time-series analysis[C]//Proceedings of the 34th international ACM SIGIR conference on Research and Development in Information Retrieval. Beijing, China, 2011: 1171-1172.

[16] Khodaei A, Alonso O. Temporally-aware signals for social search[C]//SIGIR 2012 Workshop on Time-aware Information Access.Portland, USA, 2012:45-49.

[17] Bannur S, Alonso O. Analyzing temporal characteristics of check-in data[C]//Proceedings of the 23rd International Conference on World Wide Web.Seoul, Republic of Korea, 2014: 827-832.

[18] Gupta D, Berberich K. Temporal query classification at different granularities[C]//International Symposium on String Processing and Information Retrieval. Springer, Heidelberg, 2015: 156-164.

[19] Joho H, Jatowt A, Blanco R. NTCIR temporalia: a test collection for temporal information access research[C]// Proceedings of the 23rd International Conference on World Wide Web. Seoul, Republic of Korea, 2014:

845-850.

[20] Zhao Y, Hauff C. Temporal Query Intent Disambiguation using ime-eries ata[C]//Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. Pisa, Italy, 2016: 1017-1020.

[21] Efron M, Golovchinsky G. Estimation methods for ranking recent information[C]//Proceedings of the 34th international ACM SIGIR Conference on Research and Development in Information Retrieval. Beijing, China, 2011: 495-504.

[22] Kanhabua N, Nejdl W. On the value of temporal anchor texts in wikipedia[C]//SIGIR 2014 Workshop on Temporal, Social and Spatiallyaware Information Access. Queensland, Australia, 2014:122-131.

[23] Kulkarni A, Teevan J, Svore K M, et al. Understanding temporal query dynamics[C]//Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. Hong Kong, China, 2011: 167-176.

[24] 张晓娟,陆伟,周红霞.用户查询中潜在时间意图分析及其检索建模[J].现代图书情报技术,2011(11):38-43.(Zhang Xiaojuan, Lu Wei, Zhou Hongxia. Analyzing and retrieval modeling on implicit temporal intents in users' queries[J]. New Technology of Library and Information Service, 2011(11):38-43.)

[25] Kanhabua N, Nørvåg K. Improving temporal language models for determining time of non-timestamped documents [C]//International Conference on Theory and Practice of Digital Libraries. Springer, Heidelberg, 2008: 358-370.

[26] Zhao X, Jin P, Yue L. Discovering topic time from Web news[J]. Information Processing & Management, 2015, 51(6): 869-890.

[27] Rahoman M M, Ichise R. A proposal of a temporal semantics aware Linked Data information retrieval framework [J]. Journal of Intelligent Information Systems, 2017(11):1-23.

[28] Cheng S,Arvanitis A,Hristidis V. How fresh do you want your search results? [C]//Proceedings of the CIKM'13. California, USA, 2013:1271-1280.

[29] Diaz F. Integration of news content into Web results[C]//Proceedings of the WSDM'09. Barcelona, Spain, 2009: 182-191.

[30] Konig A C, Gamon M, Wu Q. Click-through prediction for news queries[C]//Proceedings of the SIGIR'09. Boston, USA, 2009:347-354.

[31] Boer V, Someren M, Wielinga B J. Extracting historical time periods from the Web[J]. Journal of the American Society for Information Science and Technology, 2010, 61(9): 1888-1908.

[32] Jatowt A, Yeung C. Extracting collective expectations about the future from large text collections [C]// Proceedings of the 20th ACM International Conference on Information and Knowledge Management. Glasgow, UK, 2011: 1259-1264.

[33] Maio C, Fenza G, Gallo M, et al. Time-aware adaptive tweets ranking through deep learning[J]. Future Generation Computer Systems, 2017,34(11):37-48.

[34] Toyoda M, Kitsuregawa M. What's really new on the Web?: identifying new pages from a series of unstable Web snapshots[C]//Proceedings of the 15th international conference on World Wide Web. Gold Coast, Australi, 2006: 233-241.

[35] Kanhabua N, Nørvåg K. Exploiting time-based synonyms in searching document archives[C]//Proceedings of the

10th Annual Joint Conference on Digital Libraries. Edinburgh, Uk, 2010: 79-88.

[36] Jatowt A, Kawai Y, Tanaka K. Detecting age of page content[C]//Proceedings of the 9th Annual ACM International Workshop on Web Information and Data Management. McLean, USA, 2007: 137-144.

[37] Garcia-Fernandez A, Ligozat A L, Dinarelli M, et al. When was it written? Automatically determining publication dates[C]//International Symposium on String Processing and Information Retrieval. Springer, Heidelberg, 2011: 221-236.

[38] Fukumoto F, Suzuki Y. Learning timeline difference for text categorization[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 2015:98-104.

[39] Peetz M H, Meij E, Rijke M. Using temporal bursts for query modeling[J]. Information Retrieval, 2014, 17(1): 74-108.

[40] Ganesh J, Gupta M, Varma V. Improving tweet representations using temporal and user context[C]// European Conference on Information Retrieval. Springer, Cham, 2017:575-581.

[41] Jatowt A, Yeung C M A, Tanaka K. Generic method for detecting focus time of documents[J]. Information Processing & Management, 2015, 51(6): 851-868.

[42] Jatowt A, Yeung C M A. Studying how the past is remembered: towards computational history through large scale text mining[C]//ACM International Conference on Information and Knowledge Management. ACM, 2011: 1231-1240.

[43] Jatowt A, Yeung C M A, Tanaka K. Estimating document focus time[C]//ACM International Conference on Conference on Information & Knowledge Management. ACM, 2013:2273-2278.

[44] Jatowt A, Antoine É, Kawai Y, et al. Mapping temporal horizons: analysis of collective future and past related attention in Twitter[C]//Proceedings of the 24th International Conference on World Wide Web. Florence, Italy, 2015: 484-494.

[45] Abujabal A, Berberich K. Important events in the past, present, and future[C]//Proceedings of the 24th International Conference on World Wide Web. Florence, Italy, 2015: 1315-1320.

[46] Berberich K, Bedathur S J, Sozio M, et al. Bridging the terminology gap in web archive search[C]//12th International Workshop on the Web and Databases. Rhode Island, USA, 2009:47-54.

[47] Tahmasebi N, Gossen G, Kanhabua N, et al. Neer: an unsupervised method for named entity evolution recognition[C]//Proceedings the 24th International Conference on Computational Linguistics. Mumbai, India, 2012: 65-70.

[48] Kaluarachchi A C, Varde A S, Bedathur S, et al. Incorporating terminology evolution for query translation in text retrieval with association rules[C]//Proceedings of the 19th ACM International Conference on Information and knowledge management. Toronto, Canada, 2010: 1789-1792.

[49] Zhang Y, Jatowt A, Bhowmick S S, et al. Omnia mutantur, mihil interit: connecting past with present by finding corresponding terms across time[C]//Proceedings of Conference of the Association for Computational Linguistics. Beijing, China, 2015:645-655.

[50] Berberich K, Bedathur S J, Sozio M, et al. Bridging the terminology gap in Web archive search[C]//12th International Workshop on the Web and Databases, Rhode Island, USA, June 28, 2009.

[51] Radinsky K, Agichtein E, Gabrilovich E, et al. A word at a time: computing word relatedness using temporal se-

mantic analysis[C]//Proceedings of the 20th international conference on World Wide Web. Hyderabad, India, 2011: 337-346.

[52] Chundi P, Subramaniam M, Weerakoon R M A. Extracting temporal equivalence relationships among keywords from time-stamped documents[C]//International Conference on Database and Expert Systems Applications. Toulouse, France, 2011: 110-124.

[53] Wei Z, Zhou L, Li B, et al. Exploring tweets normalization and query time sensitivity for twitter search[C]// The 20th Text REtrival Conference (TREC 2011). Geithersburg, MD, USA, 2011:60-65.

[54] Setty V, Bedathur S, Berberich K, et al. InZeit: efficiently identifying insightful time points[J]. Proceedings of the VLDB Endowment, 2010, 3(1-2): 1605-1608.

[55] Dong A, Zhang R, Kolari P, et al. Time is of the essence: improving recency ranking using twitter data[C]// Proceedings of the 19th international conference on World Wide Web. Raleigh, USA , 2010: 331-340.

[56] Inagaki Y, Sadagopan N, Dupret G, et al. Session based click features for recency ranking[C]//AAAI. Georgia, USA, 2010: 1334-1339.

[57] Chang P T, Huang Y C, Yang C L, et al. Learning-based time-sensitive re-ranking for web search[C]//Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. Portland, Oregon, USA, 2012: 1101-1102.

[58] Zhang R, Chang Y, Zheng Z, et al. Search result re-ranking by feedback control adjustment for time-sensitive query[C]//Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics.Boulder, Colorado , 2009: 165-168.

[59] Meng Y, Rumshisky A, Romanov A. Temporal information extraction for question answering using syntactic dependencies in an lSTM-based architecture[J]. arXiv preprint arXiv:1703.05851, 2017.

[60] 王文,王树锋,朱锡芳.一种基于舆情敏感新信息的发现和搜索方法[J].微电子学与计算机,2012,30(12).113-116.(Wang Wen, Wang Shufeng, Zhu xifang. A method of discovering and searching sensitive recency information based on public opinion[J]. Microelectronics & Computer, 2013, 30(12): 113-116.)

[61] 李筱文.基于时态语义的 Web 信息检索关键技术研究[D].合肥:中国科学技术大学,2011.(Li Xiaowen, Exploiting key issues on Temporal Web information retrieval[D]. Hefei: University of Science and Technology of China, 2011.)

[62] Baeza-Yates R, Saint-Jean F, Castillo C. Web structure, dynamics and page quality[C]//International Symposium on String Processing and Information Retrieval. London, UK, 2002: 117-130.

[63] Berberich K, Vazirgiannis M, Weikum G. Time-aware authority ranking[J]. Internet Mathematics, 2005, 2(3): 301-332.

[64] Cho J, Roy S, Adams R E. Page quality: in search of an unbiased Web ranking[C]//Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data. Baltimore, USA, 2005: 551-562.

[65] Dai N, Davison B D. Freshness matters: in flowers, food, and web authority[C]//Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Geneva, Switzerland, 2010: 114-121.

[66] Strötgen J, Gertz M. Proximity 2-aware ranking for textual, temporal, and geographic queries[C]//Proceedings of the 22nd ACM International Conference on Information & Knowledge Management. San Francisco, USA, 2013:

739-744.

[67] Kanhabua N, Blanco R, Matthews M. Ranking related news predictions[C]//Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. Beijing, China, 2011: 755-764.

[68] Campos R, Dias G, Jorge A M, et al. Enriching temporal query understanding through date identification: how to tag implicit temporal queries?[C]//Proceedings of the 2nd Temporal Web Analytics Workshop. Lyon, France, 2012: 41-48.

[69] Gupta D, Berberich K. Diversifying search results using time[C]//European Conference on Information Retrieval. Padua, Italy, 2016: 789-795.

[70] Li X, Liu B, Philip S Y. Time sensitive ranking with application to publication search[M]. Springer International Publishing, 2010: 187-209.

[71] Dakka W, Gravano L, Ipeirotis P. Answering general time-sensitive queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(2): 220-235.

[72] Zahedi M S, Aleahmad A, Rahgozar M, et al. Time sensitive blog retrieval using temporal properties of queries [J]. Journal of Information Science, 2017, 43(1):103-121.

[73] 卫冰洁, 史亮, 王斌. 一种融合聚类和时间信息的微博排序新方法[J]. 中文信息学报, 2015, 29(3): 177-183,189.(Wei Bingjie, Shi Liang, Wang Bin. Combining cluster and temporal information for Microblog search[J]. Journal of Chinese Information Processing, 2015, 29(3):(177-183,189.))

[74] 卫冰洁, 王斌. 面向微博搜索的时间感知的混合语言模型[J]. 计算机学报, 2014, 37(1): 229-237.(Wie Bingjie, Wang Bin, Time-aware mixed language model for Micoblog search[J]. Chinese Journal of Computers, 2014,37(1):229-237.)

[75] Lin S, Jin P, Zhao X, et al. Exploiting temporal information in Web search[J]. Expert Systems with Applications, 2014, 41(2): 331-341.

[76] Vicente-Diez M T, Martinez P. Temporal semantics extraction for improving web search[C]//2009 20th International Workshop on Database and Expert Systems Application. Linz, Austria, 2009: 69-73.

[77] Arikan I, Bedathur S, Berberich K. Time will tell: leveraging temporal expressions in ir[C]// Proceedings of the WSDM '09 Workshop on Exploiting Semantic Annotations in Information Retrieval. Barcelona, Spain, 2009: 67-73.

[78] Berberich K, Bedathur S, Alonso O, et al. A language modeling approach for temporal information needs[C]// European Conference on Information Retrieval. Milton Keynes, UK, 2010: 13-25.

[79] Efron M. Query-specific recency ranking: survival analysis for improved microblog retrieval[J]. Proceedings of the TAIA, 2012, 12.

[80] Brucato M, Montesi D. Metric spaces for temporal information retrieval[C]//European Conference on Information Retrieval. Amsterdam, The Netherlands, 2014: 385-397.

[81] 林盛, 金培权, 赵旭剑,等. 时间感知的 Web 搜索研究[J]. 计算机学报, 2015(11):2163-2171.(Lin Sheng, Jin Peiquan, Zhao Xujian, et al. Research on time-aware Web search[J].Chinese Journal of Computers, 2015, 38(11):2163-2171.)

[82] Alonso O, Gertz M, Baeza-Yates R. Enhancing document snippets using temporal information[C]//International

Symposium on String Processing and Information Retrieval. Ouro Preto, Brazil 2011: 26-31.

[83] Svore K M, Teevan J, Dumais S T, et al. Creating temporally dynamic Web search snippets[C]//Proceedings of the 35th international ACM SIGIR Conference on Research and Development in Information Retrieval. Portland, USA, 2012: 1045-1046.

[84] Alonso O, Gertz M. Clustering of search results using temporal attributes[C]//Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, USA, 2006: 597-598.

[85] Alonso O, Gertz M, Baeza-Yates R. Clustering and exploring search results using timeline constructions[C]//Proceedings of the 18th ACM Conference on Information and Knowledge Management. Hong Kong, China, 2009: 97-106.

[86] Campos R, Dias G, Jorge A M, et al. GTE-cluster: a temporal search interface for implicit temporal queries [C]//European Conference on Information Retrieval. Amsterdam, The Netherlands, 2014: 775-779.

[87] Berberich K, Bedathur S. Temporal diversification of search results[C]//Proceedings of the SIGIR 2013 Workshop on Time-aware Information Access(TALA 2013). Dublin, Ireland ,2013:45-51.

[88] Whiting S, Jose J M, Alonso O. Temporal dynamics of ambiguous queries[C]//Proceedings of the SIGIR 2015 Workshop on Time-aware Information Access (TAIA2015). Santiago, Chile. 2015:92-97.

[89] Aigner W, Miksch S, Müller W, et al. Visualizing time-oriented data—a systematic view[J]. Computers & Graphics, 2007, 31(3): 401-409.

[90] Odijk D, Santucci G, Rijke M, et al. Time-aware exploratory search: exploring word meaning through time [C]//Proceedings of the TAIA'12 Workshop in conjunction with SIGIR'12. Portland ,USA, 2012:18-23.

[91] Tran N K, Ceroni A, Kanhabua N, et al. Back to the past: supporting interpretations of forgotten stories by time-aware re-contextualization[C]//Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. Shanghai, China, 2015: 339-348.

沈 思 南京理工大学经济管理学院信息管理系讲师。江苏 南京 210094。
李成名 南京师范大学文学院硕士研究生。江苏 南京 210000。
吴 鹏 南京理工大学经济管理学院信息管理系教授,博士生导师。江苏 南京 210094。

(收稿日期:2017-10-12;修回日期:2018-03-20)