

# 信息聚合概念的构成与聚合模式研究\*

曹树金 马翠嫦

**摘 要** 信息聚合正成为信息组织与检索领域的研究热点。在对现有信息聚合相关概念进行阐释的基础上,本研究提出以情景、关系和对象作为信息聚合概念的主要构成,并依此对当前国内外信息聚合的典型应用和研究前沿进行梳理,依据聚合的情景、关系和信息粒度大小,将现有研究和实践中相关聚合模式归纳为情景聚合、语义聚合、引用聚合、社会网络聚合和粒度聚合五种主要模式和十二种表现形式;在此基础上提出多模式综合的深度聚合发展趋势;情景是决定关系和粒度等其他聚合要素的依据,情景的细化和深入研究势必带来多模式的综合应用,概念和实体、引用和社会网络等关系正被综合运用,而对不同粒度信息对象的聚合则是情景和关系聚合的落脚点。图1。参考文献96。

**关键词** 信息聚合 概念构成 信息单元 聚合模式

**分类号** G254

## Information Aggregation: Concept and Paradigms

CAO Shujin & MA Cuichang

### ABSTRACT

Aiming at bringing the dispersive information into a universal interface according to users' needs, information aggregation is becoming a new realm in the field of information organization and information retrieval. Focusing on the concept of information aggregation as the starting point, this paper proposes the core elements of information aggregation as the conceptual framework to classify the typical cases of information aggregation researches and practices into several paradigms. The cases related to information aggregation are collected from the following sources: the research projects from the iSchools worldwide, the international conference papers in the field of LIS from 2009 to 2013, the research papers from five LIS international journals (including: *Journal of the American Society for Information Science and Technology*, *Journal of Documentation*, *Journal of Information Science*, *Information Processing and Management*, *Scientometrics*) from 2009 to 2013, the research projects sponsored by both the National Natural Science Foundation and the National Social Science Fund in China, and the thesis or dissertations from CNKI. Within this range, this pa-

\* 本文系国家社会科学基金青年项目“语义和情景关联的网络资源聚合单元分类体系构建与评估研究”(编号:14CTQ015)和国家社会科学基金重大项目“基于特定领域的网络资源知识组织与导航机制研究”(编号:12&ZD222)的研究成果之一。(This article is an outcome of the youth project “Development and Evaluation of Taxonomy of Aggregation Units for Academic Web Resources According to Semantic and Contextual Relationship” (No. 14CTQ015) and the project “Research on Knowledge Organization and Navigation Mechanism for Web Resources in Specific Domain” (No. 12&ZD222) supported by National Social Science Foundation of China.)

通信作者:马翠嫦,Email: macch@mail.sysu.edu.cn,ORCID:0000-0002-2478-4714 (Correspondence should be addressed to MA Cuichang,Email: macch@mail.sysu.edu.cn,ORCID: 0000-0002-2478-4714)

per selects the typical cases according to data integrity and representativeness.

After clarifying the concepts of information aggregation, this paper proposes the core elements of information aggregation as context, relationship and information objects. The main paradigms of information aggregation are classified as contextual aggregation, semantic aggregation, citation aggregation, social network aggregation, fine-grained aggregation; their manifestations include aggregation of situational information for events, aggregation relying on situational information, personal information aggregation relying on user models, aggregation relying on semantic networks, aggregation relying on citation or web links, aggregation relying on social network, aggregation based on information units, aggregation based on information fusion and aggregation based on knowledge units. Eventually, this paper proposes the development trend of integration of the five paradigms, i.e. “in-depth aggregation”. Among the core elements of information aggregation, the context of aggregation would serve as the basis of the other two elements, and the development of which would lead to the integratively application of multi-paradigms; the relationships for aggregation relying on concept and entities, citation and social network are utilized integratively in information aggregation; multi-granularity aggregation of information objects would be the result of aggregation relying on context and relationship.

The originality of this paper lies in that it proposes the conceptual framework of information aggregation through clarification, interpretation and expansion of existing concepts or definitions, the main paradigms of information aggregation and their manifestations based on this conceptual framework; it further proposes the development trend of in-depth aggregation with multi-paradigm integration. 1 fig. 96 refs.

## KEY WORDS

Information aggregation. Conceptual framework. Information units. Aggregation paradigm.

## 0 引言

网络和数字化环境下,海量信息资源为用户带来丰富数据和信息的同时,也给用户带来了信息超载的难题。因此,信息组织和检索在解决海量网络信息开发利用的同时,还肩负着搭建从信息资源到用户多样化信息需求之间桥梁的重任。

然而,以往信息组织的研究和实践偏重信息资源的开发利用,而真实情境下用户的复杂信息需求在主流信息组织和检索模式中未得到充分的考虑和支持。例如,在当前典型的信息检索系统中,用户往往需要对来源分散的信息进行逐一检索和相关性判别,并对结果进行人工整合,才能解决一个具体的问题,这为时间紧迫的用户带来了负担<sup>[1]</sup>。因而,用户需要能帮

助他们按照需求从多个分散来源中获取信息并以整合方式呈现的信息服务方式——信息聚合由此而生。

简言之,信息聚合是指根据用户的复杂信息需求,对来源分散的多类型信息资源进行采集、筛选、组织、整合和呈现的信息组织与检索方式。聚合搜索(Aggregated Search)的概念在信息科学国际会议 ACM SIGIR 2008 中被提出时,就被认为是信息检索的新范式<sup>[2]</sup>与新趋势<sup>[3]</sup>。在国内,国家社会科学基金和国家自然科学基金立项支持的与“聚合”有关的项目至今累计达十余个。可见,信息聚合作为信息组织与检索研究的新问题正得到包括图书情报学在内的诸多领域的关注。适时厘清信息聚合的内涵与概念构成,明晰其发展历史和已有基础,进而对现有的聚合理论研究与实践模式进行梳理,可为聚合研究提供理论基础和发展的路向。

## 1 信息聚合概念的构成

《牛津现代英汉双解字典》中,“聚合”一词与英文“Aggregation”对应,是指异类成分组成的集合,或由很多细小单元组成的集合组成一个整体<sup>[4]</sup>。杜晖指出,聚合的概念来源于化学领域,是描述高分子化学领域中单体小分子通过相互连接成为新的高分子化合物的专用术语<sup>[5]</sup>。可见,信息组织领域中的聚合概念应具有多来源信息单元的融合和重组的含义,指多源信息单元的聚合,包含聚合对象粒度这一要素。

北卡罗莱纳大学信息与图书馆学院信息检索领域的研究人员 Arguello 等提出了内容聚合的概念,认为内容聚合是指特定情境下围绕核心内容(Core Content)和垂直内容(Vertical Content)进行信息筛选、组织、排序与呈现的任务与方法<sup>[6]</sup>。与前述的聚合对象粒度这一要素相比,内容聚合的概念关注的多源异构的信息“内容”实际上与“聚合对象”对应,同时还增加了“情景”要素。情景是由宏观社会环境下的信息组织和检索环境以及用户需求共同决定的。作为聚合对象的信息内容的粒度,主要根据宏观社会环境下用户的复杂信息需求而定,同时还受到来源信息粒度及信息组织和检索系统性能的影响。情景和聚合对象两个要素之间衍生出第三要素,即涵盖聚合对象与用户之间关系以及聚合对象之间关系的“关系”要素。因而,信息组织领域中的信息聚合可视为面向用户需求的、基于多维聚合关系的、多来源、多粒度信息对象的筛选、组织、排序与呈现,包括聚合情景、聚合对象粒度、聚合对象之间关系及用户和聚合对象之间关系这三大构成。

与信息聚合相近的概念还有信息整合、信息联合(Syndication)、信息融合和信息集成等。此外,网络聚合搜索通常被等同于统一搜索(Unified Search)、混合垂直搜索(Blended Vertical Search)或通用搜索(Universal Search)<sup>[3]</sup>等概

念。正如本文开篇所提到的,信息聚合与这些概念的区别在于,这些概念主要从信息资源开发利用的角度提出,忽略了聚合所关注的“情景”要素以及由情景要素所带来的聚合对象、聚合对象之间关系和聚合对象与用户之间关系的差异。

可见,信息聚合是在当前搜索系统并未满足用户复杂多样的信息需求的情况下,以“情景”因素的融入为主要特征,在相应技术和理论基础之上发展而来的信息组织与搜索模式的革新。内容聚合的概念可包括情景、聚合对象和关系三大要素,其中情景决定聚合的领域、目标、资源特征、用户特征、工具和方法,是聚合模式的决定因素;信息是聚合控制的基本对象,面向信息融合和知识发现的各种应用是聚合结果;关系是聚合的基本依据,包括情景、聚合对象和用户三者本身蕴含的关系以及三者之间的关系。

## 2 聚合模式分析方法

虽然近年来信息聚合问题成为信息组织与检索研究的新热点,但 Arguello 认为内容聚合的现象可追溯到更传统的表现形式,如:报纸可视为最早的“聚合媒体”,数字环境下聚合系统则表现为内容过滤系统,万维网初期更多呈现为主题导航式的手工的内容聚合、网络内容的自动聚合和新闻推荐等形式。当前技术环境下,内容聚合者则采用复合的、实时的、基于地理空间信号、语义分析和社会信息分析的方式进行<sup>[6]</sup>。国内对于信息聚合的研究也涵盖了馆藏资源聚合和网络资源聚合等方面,如,伍革新和程秀峰提出数字图书馆语义检索从功能上可以划分为面向服务的聚合与检索、面向用户的聚合与检索、面向知识推理与知识创新的聚合与检索<sup>[7]</sup>。可见,信息聚合具有多样化的应用领域和应用模式,加上学界对于信息聚合概念的认识和聚合类型划分尚不一致,当前与聚合模式相关的研究和实践是多样而又分散的。

本研究以信息聚合的概念构成为基础,对当前国内外关于信息聚合的典型应用和研究前沿进行梳理,归纳出聚合的主要模式。为获取国内外图情领域关于信息聚合研究与实践的信息,本研究搜集并整理了以下项目和资料:全球 iSchool 图书情报学院的教师科研项目、2009—2013 年间图书情报领域的国际会议论文、2009—2013 年间图情领域的五种国际期刊的论文(*Journal of the American Society for Information Science and Technology*, *Journal of Documentation*, *Journal of Information Science*, *Information Processing and Management*, *Scientometrics*)、我国国家自然科学基金和国家社会科学基金资助项目论文和 CNKI 中的学位论文,并以此为线索进行扩展。在此范围内,选取资料完整的、有代表性的研究和实践案例进行分析,对聚合模式进行归纳和梳理。

通过分析发现,这些项目和研究中包含的

聚合模式主要是以情景、对象、关系中的一种或几种为依据进行的信息聚合,涵盖特定情境下多类型信息资源的采集、抽取、组织、分析、应用等多个方面的内容。围绕这些特征,我们提出本研究模式分析的思路,即:以信息聚合的主要构成要素为模式分析的立足点,围绕情景、关系和聚合对象粒度三种主要的聚合依据,按照其研究和实践的侧重点不同,划分为情景聚合、语义聚合、引用聚合、社会网络聚合和粒度聚合五种主要模式。以此为框架,以聚合情景分析为基础,以聚合机制解释和聚合应用介绍为主要内容,对现阶段信息聚合研究和实践的基本模式及其具体表现形式进行梳理,具体如图 1 所示。这三种聚合模式并非相互排斥、独立存在,而是互相包容和渗透。值得注意的是,以下各模式下所讨论的研究和实践例子除了具备该模式的显著特征外,也可能同时具有其他模式的特征和要素,甚至是其中几种模式的复合体。

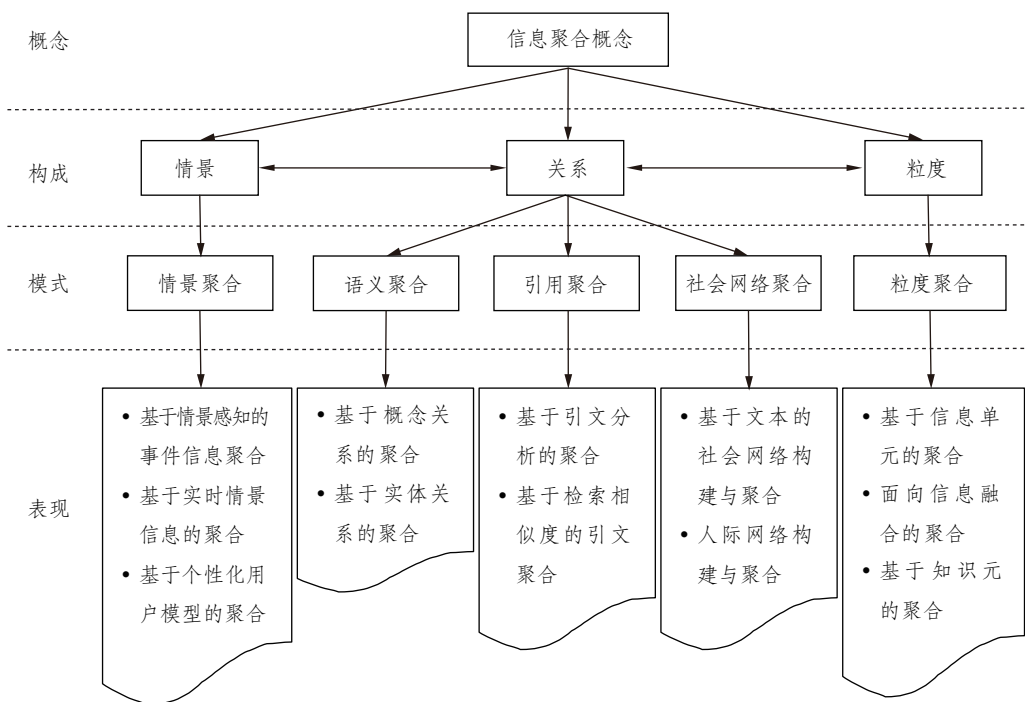


图 1 基于信息聚合概念构成的聚合模式与表现形式

### 3 信息聚合的主要模式与表现

#### 3.1 基于情景的聚合

情景是指领域、用户、时间、地点等可影响聚合目标、聚合对象特征、聚合方法和工具的应用的社会环境因素。情景聚合模式即基于应用目标、用户需求、行为或时间地点等客观情景因素进行的信息对象的采选、整合与动态呈现。该模式可分为基于情景感知的事件信息聚合、基于实时信息的新闻或消息聚合和基于个性化用户模型的聚合三种表现形式。

##### 3.1.1 基于情景感知的事件信息聚合

大数据环境下,科研、商业、政府管理、军事和社会生活中使用的各类传感器产生的多媒体信息是海量信息的主要组成之一。基于情景感知的事件信息聚合模式是指通过对来自多个传感器的实时、海量的多媒体信息进行事件情景识别、事件相关信息或数据抽取和事件整合的聚合模式。基于大规模多媒体传感器的应用系统被称为情景感知的应用<sup>[8]</sup>。

相关研究和实践包括:Ramachandran以视频监控为例构建情景感知的应用模型,在监测器识别和异常事件触发的基础上,通过多传感器和追踪器进行事件信息(如:位置等)的持续聚合和联合追踪,形成面向实际应用的相关知识<sup>[8]</sup>。Gao等构建了一个情景信息监测的系统,通过异构信息流聚合方法提供个性化信息和行动提醒。该系统通过统一的数据模型聚合异构网络数据流,并监测个人或整合数据流的时间和空间特性,从而对复杂情景进行监测并提供个性化信息服务与提醒<sup>[9]</sup>。Dao等构建涵盖网络、实体和社会系统的事件信息管理平台,以帮助领域专家实现大规模动态事件数据的监测和实时情景预测。该平台设计半自动的情景识别模型,通过输入来自传感器的或书面报告的、实时的或档案信息中的同构或异构数据格式,对数据流实现浅层或深层的混合式整合,从而产生可视化的分析方法或行动和警告,并帮助发

现隐藏的时间模式<sup>[10]</sup>。

##### 3.1.2 基于实时情景信息的聚合

实时情景信息主要是指用户所处的地理位置、时间甚至是社会关系网络等即时的或历时的情景信息。基于实时信息的聚合是当前聚合研究的热点。

相关研究和实践包括:Wenze和Kießling开展首个对多来源的位置—社会网络数据进行聚合、建模和分析的研究,从Facebook、Google+、Twitter和Foursquare中抽取数据并生成物化的空间架构,在此基础上生成扩展的用户模型并实现基于用户模型的相似性分析,从而应用到用户间的场景式推荐中<sup>[11]</sup>。Singh等在研究中提出,微博等社会网络程序产生的空间—时间—主题数据的整合,可帮助检测事件,并理解同一事件在不同时间—空间粒度中发展的情景。作者受到传统图像像素代表某位置中光子强弱的聚合的启发,将处于不同地理位置的用户兴趣层次作为社会聚合的像素,将这种社会像素以时间—空间的方式联系起来即可形成社会图像和音频。作者提出通过情景监测操作器实现基于时间、空间和主题社会像素的图像和音频聚合<sup>[12]</sup>。Liu提出的LocalSavvy原型系统,将不同地点对于相同新闻事件的官方与非官方等不同观点聚合到统一界面,同时进行摘要呈现,用户可对不同的观点进行比较,从而丰富用户新闻阅读体验<sup>[13]</sup>。

##### 3.1.3 基于个性化用户模型的聚合

基于个性化用户模型的聚合可分为面向用户个体的建模与聚合和面向用户群体的建模与聚合。其中,面向个体的用户模型聚合可划分为个性化新闻推荐中的用户建模(如:主题跟踪和语义标签匹配模型<sup>[14]</sup>、基于用户交互行为特征的多变量优化模型<sup>[14]</sup>、动态的行为学习模型<sup>[15]</sup>等)和场景模型构建(如:基于生活事件法的电子政务服务场景模型<sup>[16-17]</sup>)。

基于多来源数据的用户群组建模研究近年来备受关注,下面介绍较为典型的研究和实践。ASTERIX(Advances in Small Trials dEsign for

Regulatory Innovation and eXcellence, 创新卓越小型实验系统) 大数据信息管理系统, 旨在为洛杉矶地区产生的地理、商业、人口、新闻、本地相关或提及本地的博客和微博等多源、异构和动态的事件数据管理提供基本框架, 构建动态数据模型, 最终实现基于用户兴趣群组(如选举、电影等)的事件信息聚合<sup>[18-19]</sup>。微软亚洲研究院的研究人员 Zhuang 等提出移动环境下基于潜在情境因素的感知与个性化商业机构推荐问题。在对手机点击行为进行大规模分析的基础上识别商业机构类型和条目, 通过用户的特定情境估算相关商业机构的概率, 推荐与用户情景(过往行为)和感知情景(时间和地点)相关的机构类型和机构。该研究在 Windows Phone 7 设备的基础上开发个性化推荐应用<sup>[20]</sup>。

### 3.2 基于关系的聚合

基于关系的聚合是指特定情境下依据聚合对象之间的各类关系进行聚合, 主要包括基于语义关系的聚合、基于引用关系的聚合和基于社会关系的聚合。

#### 3.2.1 语义聚合模式

语义聚合模式着重探索文本信息资源内容所包含的概念间或实体间关系, 从而通过语义关系网络实现多类型资源(文本、馆藏、数据和服务)的聚合。

概念关系可分不同的层次, 主要包括概念的包含关系, 如分类法的类目等级关系, 叙词法的属分关系, 本体类的树状结构包含关系等。Peace 构建了关于医护实践知识的知识组织系统, 辅助计算机理解和分析<sup>[21]</sup>; 李劲等研究如何揭示馆藏资源内部的知识对象以及知识对象间的语义关系, 通过聚类、融合和重组使分散无序、相对独立的馆藏资源重新组织为一个新的有机整体<sup>[22]</sup>。

实体关系, 如人物、时间、地点、事件等实体是和内容主题并存的另一大类关系载体, 在基于非主题因素的关系的揭示与发现方面发挥着不可忽视的作用。基于实体关系的聚合可包

括: 面向馆藏资源实体关系的聚合(如: 基于 FRBR 模型的 OPAC 系统聚合<sup>[23]</sup>), 基于关联数据的馆藏信息资源聚合<sup>[24]</sup>), 面向事务实体关系的聚合(如: 基于语义网络的多来源水质量管理数据源的聚合<sup>[25]</sup>), 生物等学科领域的数据与服务聚合<sup>[26-27]</sup>), 以及面向具体领域实体关系的聚合(如: ad hoc 网络层面的基于语义和规则网络的软件构件聚合<sup>[28]</sup>)。

#### 3.2.2 引用聚合模式

信息资源间存在大量的链接和引用关系, 这些关系构成了特定研究主题的知识网络, 其单向无回路的特征揭示了学科主题的知识结构和发展过程<sup>[29]</sup>, 被认为是获得某领域文献知识结构的方法<sup>[30]</sup>。基于引用关系的聚合着重探索资源之间的链接、参考和引证关系, 从而为聚合提供依据。相关研究集中在基于实体间引用关系的聚合、数字图书馆的馆藏资源聚合与可视化研究方面, 包括基于文献间引用关系的聚合、基于实体间引用关系的聚合和基于检索相似度的引文聚合。

邱均平带领的团队对基于引文分析的聚合进行系列研究, 如: 基于阈值和权值的数字期刊论文聚合<sup>[29]</sup>, 基于文献耦合关系与作者耦合关系的信息资源聚合<sup>[5]</sup>, 基于共现与耦合的馆藏资源聚合<sup>[31]</sup>等。此外, 曾建勋提出利用学术文献引证关系和 WWW 链接机制构造中文知识链接门户的思想, 将系统数据库结构分为来源文献库、被引文献库、作者库、基金库、期刊载文表、期刊引文表等, 通过“来源文献唯一标识”将各数据库相关记录联系起来, 通过聚合维度数据规范、优化检索关联字典等提高链接和统计的几率和效率<sup>[32]</sup>。

Ding 等提出基于文献内容单元的引用分析, 从语义和语法的层面阐释引用的价值, 及其在引用动机分类、引用摘要、信息检索、推荐和预测及知识图谱挖掘等方面的应用<sup>[33]</sup>。在基于实体方法的引用分析中, 知识实体被定义为关键词、论题、主题类目、数据集、关键方法、关键理论、领域实体等知识的载体, 其在科学交流过

程中所体现的关系可为知识聚合提供依据。按照这一方法,作者在生物医学数据库 PubMed 中对糖尿病药物二甲双胍相关实体进行引用分析,从而形成基于生物实体的引用网络,为相关药物研发中的知识发现提供支持<sup>[34]</sup>。

基于检索相似度的引文聚合方面,Strotmann 和 Zhao 提出对引文数据集合的相似度矩阵进行多因素因子分析,通过对用户搜索结果和/或与其密切相关的文档进行计算,为用户提供数字图书馆文献的交互式地图。搜索结果中的每个可视化元素可为用户提供从不同方面(作者和/或子领域)进行聚合的途径,从而帮助用户认识领域知识结构和书目间亲疏关系<sup>[30]</sup>。

### 3.2.3 社会网络聚合模式

基于社会网络的聚合是指以社会关系分析的结果为依据进行的资源聚合。根据 Freeman 的定义,社会网络分析是指以计算机技术和统计学原理作为支持,以系统数据为基础,以图形语言和技术为表现,对社会网络主体之间形成的特定结构关系进行分析,探究隐藏在复杂的社会系统表面之下的特定网络模式的分析方法<sup>[35-36]</sup>。社会网络将社会行为者表达为节点,节点之间的联系表达为边。其中,边可以表达不同类型的关系,如血缘关系、合作关系、敌对关系等,而构成网络的实体关系可以延伸到表达事件中的谁(个人或组织等社会机构)、什么(任务、事件)、哪里(位置)、为什么(观念、情感、信仰)、如何(资源、信息)和何时(时间)等类中,从而实现多维度的甚至语义层面的聚合<sup>[37]</sup>。基于社会网络的聚合主要包括两种表现形式:基于文本的社会网络构建与聚合,基于文本的人际网络构建与聚合。

伊利诺伊大学图书馆与信息学院副教授 Diesner 对基于文本的语义挖掘和社会—语义网络构建与聚合开展系列研究<sup>[38]</sup>,在此基础上开发了 ConText ①“社会正义”领域的文本分析与聚合系统。该系统从元数据中或文本<sup>[39]</sup>中获取

文本数据,通过一系列文本特征(语义的、句法的、近似的或概率的)进行关系抽取,抽取社会系统结构,从而辅助该领域进行基于文本的聚合分析<sup>[40-42]</sup>。此外,刘礼芳提出对 Flickr 网站用户进行社会网络结构划分,通过基于社会网络结构的标签学习算法,实现基于社会网络语义标签的 Web 图像聚合,从而将感兴趣的好友图像和社区图像聚合到用户个人网页中<sup>[43]</sup>。

人际网络构建与聚合方面,Diesner 等学者以某危机应对组织中的即时邮件通讯语料库为对象,构建商业机构沟通网络的结构和属性,并探索不同层次员工沟通行为的属性和模式,通过发现危机下商业机构沟通的特有模式,为商业机构建设提供决策参考<sup>[44]</sup>。王雨提出综合运用社会网络分析、矩阵算法、网络图谱等工具和方法,构建基于作者互引关系、作者合作关系、作者—关键词关系的社会网络聚合模型和可视化图谱,实现数字图书馆馆藏资源聚合<sup>[35]</sup>。胡蓉和邓小昭以结构洞理论为基础,构建计算机辅助的个人人际网络分析系统,运用矩阵分析算法和网络图技术分析现实竞争环境中企业行动者的个人人际网络结构,从而辅助企业行动者利用其人际网络和结构洞优势来获取更有效的信息资源<sup>[45-46]</sup>。

## 3.3 基于聚合对象粒度的聚合

基于聚合对象粒度的聚合主要是指对文本信息资源的解构与重组,相关研究按照知识融合与发现的深度依次为基于信息单元的聚合、信息融合和基于知识元的聚合三大类型。因而,粒度聚合模式下的研究和实践除了具有聚合对象粒度较小的共同特征外,还可能同时具有情景聚合模式、关系聚合模式的特征。

### 3.3.1 基于信息单元的聚合

基于信息单元的聚合主要是对数字或网络环境下文本信息单元中包含的相关信息的抽取、检索或重组。相关研究涉及数字文档和网

① <http://context.lis.illinois.edu>

络各类信息片段的解构与重组,既包括基于逻辑单元的聚合,也包括基于语义单元的聚合。

当前,无论是数字资源还是网络资源的聚合,大都是基于逻辑单元的聚合。如期刊论文章、节、段、句和图表等逻辑单元的聚合,如伊利诺伊大学早年开展的 NSF/DARPA/NASA 数字图书馆项目(DLI)中的 DeLiver 项目<sup>[47]</sup>、CNKI 知识元搜索<sup>①</sup>和加州大学伯克利分校的 Hearst 教授开发的生物论文搜索引擎 BioText Search Engine<sup>②</sup>。此外,Wu 等将网络检索中的锚文本检索技术改进应用到图书搜索中,以书后索引作为指向书页的锚,采用概率数据记录作为检索策略的高层次摘要模型,从而通过概率逻辑摘要实现图书书页检索和定位<sup>[48]</sup>。

Ratkiewicz 和 Menczer 以 HTML 页面的 DOM 树的标签与位置为线索,通过超链接连接起来并进行片段抽取<sup>[49]</sup>。Bessai-Mechmache 提出可支持关键词自动选择的 XML 文档非冗余要素的聚合模型,在概率网络的基础上,提供文档、要素及内容之间自然表达的链接<sup>[50]</sup>。此外,Jahn 等通过 Bing 搜索引擎,获取研究人员出版物列表的 URL,通过每一个 URL 抽取相应的内容片段,从而实现研究人员出版物列表的自动发现与聚合<sup>[51]</sup>。

语义单元聚合方面,Hearst 教授开发的系统 WordSeer<sup>③</sup>可通过字、词、句层面的文本分析,为人文社会科学学者提供语义聚合平台,提供基于词间语义关系的语法检索、用语比较、语句关系可视化和结果可视化,支持基于主题的例子聚合、文本标引、内容导航等。Blake 综合采用句法和语义的方法,从生物医学领域的实证研究论文结论中自动识别和抽取科学论断,通过区分关联、比较和观察,来捕获不同层次的事实,从而减轻信息超载的负担<sup>[52]</sup>。Zhang 识别期刊文章组成(如介绍、方法、结果和讨论)中的最小语义功能单元,及其与信息利用任务(Infor-

mation Use Task)之间的关系,从而支持基于聚合单元功能组织与呈现的期刊论文导航、精读、理解和使用<sup>[53]</sup>。

Liu 提出的 Spectrum 系统提供对当前流行的新闻主题的聚合。该系统可围绕特定主题对 blogosphere 中关于某新闻的不同观点进行自动聚类,从而支持用户对于某新闻主题的不同观点进行浏览与比较。Liu 提出的另一系统 Compare&Contrast,可分析用户提交的新闻故事,建立故事情景模型,从而动态地发现与原始新闻中可对比的实体(如任务、地点、组织等),并进行案例扩展检索,通过实时动态更新、与地理位置相连、情景分析和语义分析发现社会趋势<sup>[54]</sup>。

### 3.3.2 面向信息融合的聚合

面向信息融合的聚合是指通过多源文本或多媒体片段的摘要与融合,实现面向特定需求的多类型信息融合与服务的聚合效果,可包括文本自动摘要、多媒体信息融合。

文本自动摘要常被应用于结果聚合中,包括基于文本内容的结构化摘要(如 Lin 和 Liu 的研究<sup>[55]</sup>)、面向集合的高层次总体概要和可视化三个层面的聚合<sup>[56]</sup>。Sushmita 和 Lalmas 提出通过结果文档聚类形成虚构文档回答检索提问,用户还可通过聚合摘要页的概念获得与查询相关的情景<sup>[57]</sup>。与此类似,Ou 和 Khoo 提出社会科学领域聚合检索结果页面提供基于摘要概念和结果的组织与呈现方法<sup>[58]</sup>。Kopliku 研究基于信息块的结构化摘要、多类型内容融合、定制及其可视化的问题<sup>[3]</sup>。Kaptein 和 Marx 实现由丹麦议会手稿组成的大规模半结构化数据集的聚焦搜索和结果聚合,以图表的形式进行会议结构摘要和可视化,会议内容或部分内容被压缩成云标签<sup>[59]</sup>。

多媒体信息融合方面,Jain 报告了利用海量

① <http://www.cnki.net>

② <http://biosearch.berkeley.edu>

③ <http://wordseer.berkeley.edu>

媒体数据进行故事合成的前沿研究与主要技术。在数据获取方面,作者提出通过多个传感器和多个媒体对现实世界事件进行捕获,通过数据分析、相关事件和经历的选择,实现基于这些素材的事件重现,在此基础上合成故事<sup>[60]</sup>。

### 3.3.3 基于知识元的聚合

基于知识元的聚合是以知识元理论为基础,实现基于知识单元及其语义关系的揭示与组织,从而达到按特定目标聚合的效果。基于知识元的聚合按照应用可分为学科领域的知识元聚合与舆情领域的知识元聚合两类。

学科领域知识元聚合方面,刘植惠、文庭孝、温有奎、化柏林等对此进行了系统研究,如知识元理论研究<sup>[61-64]</sup>,知识元自动识别、标引与组织<sup>[65-71]</sup>,知识元检索系统构建等<sup>[72-76]</sup>,从不同程度上为基于知识元聚合提供参考。知识元的识别与抽取常被用于构建具有一定结构和关联的知识体系,相关研究可包括:知识链接研究<sup>[77-80]</sup>,学科领域多粒度知识组织体系构建、检索与整合<sup>[81-82]</sup>,基于知识元的领域知识系统构建,如基于知识元的中药方剂知识组织系统构建<sup>[83]</sup>、语义 wiki 知识元系统<sup>[84-85]</sup>和基于知识元基础的军事情报语义网地图<sup>[86]</sup>等。这些研究通过情境下基于知识元的知识组织系统构建,实现按照学科领域需求从更小的粒度构建和揭示知识的聚合效果。

知识元理论应用到舆情领域,形成了面向具体应用的知识元理论体系。舆情领域的研究通过提取突发事件相关要素作为知识元进行标引和建模,对要素间的关系进行探索,从而对舆情事件进行预警、追溯和决策支持。大连理工大学的研究团队对此进行了系列研究,包括:舆情事件知识元通用模型构建<sup>[87-88]</sup>、面向特定情境的舆情事件知识元模型构建<sup>[88-90]</sup>、应急决策知识元模型构建<sup>[90-92]</sup>和面向历时性分析的舆情知识元模型构建<sup>[93-95]</sup>等,从而实现面向舆情管理需求的动态信息抽取和知识整合,达到基于知识元聚合的效果。

## 4 多模式综合的深度聚合发展趋势

在情景、关系和聚合对象三个聚合概念构成中,情景是由社会环境、信息组织和检索环境以及用户需求共同决定的;聚合对象的粒度大小实际上是根据社会环境和用户需求而定的,既受到来源信息粒度的影响,又受到信息组织和检索系统性能的影响;关系则是反映聚合情景和聚合对象关系的总和与落脚点;因此,情景是决定聚合其他要素的依据。情景的细化和深入研究势必带来多模式的综合应用,通过探索不同关系类型和不同粒度信息对象与各种情境下用户需求的关联关系,可以进一步满足用户复杂的、多样的和动态的需求。

对基于关系的聚合而言,这种多模式综合发展的趋势首先表现为,语义聚合所包含的基于概念关系或实体关系的聚合普遍存在于情景聚合和粒度聚合等多种模式中;其次,基于概念间的关系和实体间关系的聚合正在融合发展,这与信息检索中强调非主题检索及主题与非主题因素融合的发展趋势不谋而合;第三,引文聚合与社会网络聚合的结合,包括社会网络分析在内的网络分析方法正被整合到引文分析中<sup>[36,96]</sup>;第四,随着网络分析方法的发展,基于社会关系和引用关系的非传统知识组织方式正逐渐与语义聚合方式结合,提供多维度展示资源分布和知识结构的组织方式。

对于粒度聚合而言,作为组织对象的信息资源粒度的细化,使得面向情景的聚合控制可以跨越载体单元而深入至内容本身,从而为面向用户多种情景下的动态需求和海量信息环境下对于目标信息的获取提供有力的支持;然而,粒度聚合在聚合单元之间关系的构建、聚合单元的识别与描述、聚合搜索与呈现等方面,有别于传统的以篇章为单元的知识组织方式:对于长文本而言,粒度聚合模式下知识组织系统构建必须考虑不同粒度层级聚合单元的主题表达,聚合粒度越小,主题数量越丰富,主题间的

语义关系越明确和丰富,为聚合单元的基于主题和非主题因素的多分面描述提供可能性,进一步增加聚合搜索的语义准确性,以及多种情景下语义聚合的智能化;对于网络环境下越来越多的评论、问答、实时讨论等类型的短文本而言,粒度聚合为短文本的组织提供了有针对性的组织方式。具体应用目标和用户认知特征等具体情景构建知识组织系统,则是短文本聚合的关键。

## 5 讨论

当前,信息聚合研究正成为图书情报领域关注的热点,然而学界对于信息聚合的研究至今仍然集中在方法、技术和应用的层面,对于信息聚合理论的梳理和构建相对较少。本研究的贡献在于:①从概念的辨析、解释和扩展出发,明晰信息聚合概念的主要构成;②在信息聚合

概念构成的基础上,提出主要的聚合依据、聚合模式及其主要表现形式,从而对各模式下的聚合机制和聚合应用进行梳理;③提出多模式综合的深度聚合发展趋势,为深度聚合的方法、技术和应用的进一步发展提供理论基础。

后续相关研究一方面可将本研究提出的信息聚合构成和聚合模式作为理论框架,从图书情报领域信息组织与信息检索的角度对每种聚合模式的具体聚合机制、方法和技术进行细化研究,从而推动信息聚合理论的发展,拓宽图书情报学的研究领域;另一方面,可结合特定学科领域的特定工作任务和舆情分析等具体情景,充分利用信息聚合的最新成果,开发面向知识服务与决策支持的深度聚合系统,充分发挥信息聚合理论在提升科学研究水平、推动社会经济发展中的作用,提升图书情报学理论在解决社会实际问题中的能力。

## 参考文献

- [1] Paris C, Wan S, Thomas P. Focused and aggregated search: a perspective from natural language generation[J]. *Information Retrieval*, 2010, 13(5): 434-459.
- [2] Murdock V, Lalmas M. Workshop on aggregated search[J]. *SIGIR Forum*, 2008, 42(2): 80-83.
- [3] Kopliku A. Aggregated search: from information nuggets to aggregated documents[C]//*Proceedings of Conférence en Recherche d'Information et Applications*, 2009: 495-502.
- [4] Thompson D. 牛津现代英汉双解词典[M]. 北京: 外语教学与研究出版社, 2003: 38. (Thompson D. The concise Oxford English dictionary. Beijing: Foreign Language Teaching and Research Press, 2003: 38.)
- [5] 杜晖. 基于耦合关系的学术信息资源深度聚合研究[D]. 武汉: 武汉大学, 2013. (Du Hui, Research on the depth of the academic information resources based on the coupling relationship[D]. Wuhan: Wuhan University, 2013.)
- [6] Arguello J, Diaz F, Shokouhi M. Integrating and ranking aggregated content on the Web[EB/OL]. (2012-05-03) [2016-03-25]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.309.8948&rep=rep1&type=pdf>.
- [7] 伍革新, 程秀峰. 面向语义检索的数字图书馆资源聚合模式分析[J]. *图书馆建设*, 2013(10): 22-26. (Wu Gexin, Cheng Xiufeng. Analysis on the resource aggregation mode of digital libraries faced the semantic retrieval[J]. *Library Development*, 2013(10): 22-26.)
- [8] Ramachandran U, Hong K, Iftode L, et al. Large-scale situation awareness with camera networks and multimodal sensing[C]//*Proceedings of the IEEE*, 2012, 100(4): 878-892.
- [9] Gao M, Singh V K, Jain R. Eventshop: from heterogeneous Web streams to personalized situation detection and

- control[C]//ACM. Proceedings of the 3rd Annual ACM Web Science Conference,2012; 105-108.
- [10] Dao M S,Pongpaichet S,Jalali L,et al. A real-time complex event discovery platform for Cyber-Physical-Social systems[C]//ACM. Proceedings of International Conference on Multimedia Retrieval,2014; 201.
- [11] Wenze F, Kießling W. Aggregation and analysis of enriched spatial user models from location-based social networks[EB/OL]. [2016-03-26]. [https://www.informatik.uni-augsburg.de/lehrstuehle/dbis/db/publications/all\\_db\\_publications/2014\\_wen\\_kie\\_georich](https://www.informatik.uni-augsburg.de/lehrstuehle/dbis/db/publications/all_db_publications/2014_wen_kie_georich).
- [12] Singh V K,Gao M,Jain R. Situation detection and control using spatio-temporal analysis of microblogs[C]//ACM. Proceedings of the 19th International Conference on World Wide Web,NC,USA,2010; 1181-1182.
- [13] Liu J H. Information diversity in Web search[D]. Illinois: Northwestern University,2009.
- [14] Krakovsky M. All the news that's fit for you[J]. Communications of the ACM,2011,6(54): 20-21.
- [15] Li L,Chu W,Langford J,et al. A contextual-bandit approach to personalized news article recommendation[C]//ACM. Proceedings of the 19th International Conference on World Wide Web, North Carolina, USA, 2010; 661-670.
- [16] Wimmer M A. Integrated service modelling for online onE-stop government[J]. Electronic Markets,2002,12(3): 149-156.
- [17] 李广建,王巍巍. 基于生活事件的信息服务整合方法及其应用[J]. 图书与情报,2013(6): 32-40. (Li Guangjian,Wang Weiwei. The integration method and its application of information services based on life-events[J]. Library & Information,2013(6): 32-40.)
- [18] Bental D,Stewart R,Aylett R,et al. Exposing connections to support serendipitous discovery[R/OL]. [2016-03-25]. <http://homepages.abdn.ac.uk/k.vdeemter/pages/sipi-2012/bental.pdf>.
- [19] Behm A,Borkar V R,Carey M J,et al. Asterix: towards a scalable,semistructured data platform for evolving-world models[J]. Distributed and Parallel Databases,2011,29(3): 185-216.
- [20] Zhuang J F,Mei T M,Hoi S C H,et al. When recommendation meets mobile: contextual and personalized recommendation on the go[C]//ACM. Proceedings of the 13th International Conference on Ubiquitous Computing, 2011; 153-162.
- [21] Peace J. The use of ontology to represent nursing knowledge about family health history and facilitate automated search for clinical practice guidelines[D]. Madison: University of Wisconsin-Madison,2008
- [22] 李劲,程秀峰,宋红文,等. 基于语义的馆藏资源深度聚合模型探析[J]. 湖北民族学院学报(自然科学版), 2013(2): 212-215. (Li Jin, Cheng Xiufeng, Song Hongwen, et al. The aggregation model of library resources based on semantic methods[J]. Journal of Hubei University for Nationalities (Natural Science Edition), 2013(2): 212-215.)
- [23] 王泽贤. 基于CNMARC的FRBR化OPAC系统开发初探[J]. 现代图书情报技术, 2008(7): 81-85. (Wang Zexian. The development preliminary research of FRBRized OPAC system based on CNMARC[J]. New Technology of Library and Information Service, 2008(7): 81-85.)
- [24] 王涛. 基于关联数据的馆藏信息资源聚合研究[J]. 图书馆学刊, 2012(8): 44-46. (Wang Tao. Research on aggregation of library resources based on linked data[J]. Journal of Library Science, 2012(8): 44-46)
- [25] Chen Z,Gangopadhyay A, Holden S H, et al. Semantic integration of government data for water quality management[J]. Government Information Quarterly, 2007, 24(4): 716-735.
- [26] Palmonari M, Sala A, Maurino M, et al. Aggregated search of data and services[J]. Information Systems, 2011, 36: 134-150.

- [27] Sala A. Data and service integration-architectures and applications to real domains[D]. Italy: University of Modena and Reggio Emilia, 2010.
- [28] Karabatis G, Chen Z, Janeja VP, et al. Using semantic networks and context in search for relevant software engineering artifacts[J]. Journal on Data Semantics XIV, 2009: 74-104.
- [29] 邱均平, 董克. 引文网络中文献深度聚合方法与实证研究: 以 WoS 数据库中 XML 研究论文为例[J]. 中国图书馆学报, 2013(2): 111-120. (Qiu Junping, Dong Ke. Methods and empirical research on deep integration of literature in citation network: case study on XML research literature from WoS [J]. Journal of Library Science in China, 2013 (2): 111-120.)
- [30] Strotmann A, Zhao D. Bibliometric maps for aggregated visual browsing in digital libraries[C]//Proceedings of the SIGIR 2008 Workshop on Aggregated Search in SIGIR, 2008: 501-512.
- [31] 邱均平, 王菲菲. 基于共现与耦合的馆藏文献资源深度聚合研究探析[J]. 中国图书馆学报, 2013(3): 25-33. (Qiu Junping, Wang Feifei. An exploration of in-depth aggregation of library document resources based on co-occurrence and coupling [J]. Journal of Library Science in China, 2013 (3): 25-33.)
- [32] 曾建勋. 中文知识链接门户的构筑[J]. 情报学报, 2006, 25(1): 63-69. (Zeng Jianxun. Construction of Chinese knowledge linking portal[J]. Journal of the China Society for Scientific and Technical Information. 2006, 25(1): 63-69.)
- [33] Ding Y, Zhang G, Chambers T, et al. Content-based citation analysis: the next generation of citation analysis[J]. Journal of the Association for Information Science and Technology, 2014, 65(9): 1820-1833.
- [34] Ding Y, Song M, Han J, et al. Entitymetrics: measuring the impact of entities[J/OL]. PloS one, 2013, 8(8): e71416 [2016-03-25]. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0071416>.
- [35] 王雨. 基于社会网络分析的数字图书馆资源聚合研究[D]. 吉林: 吉林大学, 2014: 9. (Wang Yu. Research on digital library resources aggregation based on SNA[D]. Jilin: Jilin University, 2014: 9.)
- [36] Freeman L C. Centrality in social networks conceptual clarification[J]. Social Networks, 1979, 1(3): 215-239.
- [37] Diesner J, Carley K, Tambayong L. Extracting socio-cultural networks of the Sudan from open-source, large-scale text data[J]. Computational & Mathematical Organization Theory, 2012, 18: 328-339.
- [38] Diesner J. Research & projects[EB/OL]. [2015-05-19]. <http://people.lis.illinois.edu/~jdiesner/research.html>.
- [39] Diesner J. From texts to networks: detecting and managing the impact of methodological choices for extracting network data from text data[J]. Artificial Intelligence, 2013, 27(1): 75-78.
- [40] Diesner J. Network construction based on structured and unstructured text data in context [C/OL]//Conference of International Communication Association, 2015 [2015-05-20]. [http://vanatteveldt.com/wp-content/uploads/diesner\\_context.pdf](http://vanatteveldt.com/wp-content/uploads/diesner_context.pdf).
- [41] Diesner J. ConText: software for the integrated analysis of text data and network data[C/OL]//Conference of International Communication Association, 2014 [2016-03-25]. [http://people.lis.illinois.edu/~jdiesner/calls/ICA2014/Diesner\\_ICA\\_2014.pdf](http://people.lis.illinois.edu/~jdiesner/calls/ICA2014/Diesner_ICA_2014.pdf).
- [42] Diesner J. Words and networks: how reliable are network data constructed from text data?[G]//Bertino E, Matei S A. Roles, Trust, and Reputation in Social Media Knowledge Markets Computational Social Sciences. Switzerland: Springer, 2015: 81-89.
- [43] 刘礼芳. 基于社会网络的 WEB 图像语义标注与聚合[D]. 杭州: 杭州师范大学, 2011. (Liu Lifang. Web image semantic annotation and aggregation based on social network[D]. Hangzhou: Hangzhou Normal University,

- 2011.)
- [44] Diesner J, Frantz T L, Carley K M. Communication networks from the Enron email corpus[J]. Computational & Mathematical Organization Theory, 2005, 11(3): 201-228.
- [45] 胡蓉, 邓小昭. 基于结构洞理论的个人人际网络分析系统研究[J]. 情报学报, 2005, 24(4): 485-489. (Hu Rong, Deng Xiaozhao. A study on the structural holes theory-based egocentric human intelligence network analysis system[J]. Journal of the China Society for Scientific and Technical Information, 2005, 24(4): 485-489.)
- [46] 胡蓉. 基于结构洞理论的个人人际情报网络分析系统研究[D]. 重庆: 西南师范大学, 2005. (Hu Rong. A Study on the structural holes theory-based egocentric human intelligence network analysis system[D]. Chongqing: Southwest China Normal University, 2005.)
- [47] Bishop A P. Digital libraries and knowledge disaggregation: the use of journal article components[C]//Proceedings of the Third ACM Conference on Digital Libraries, 1998: 29-39.
- [48] Wu H, Kazai G, Roelleke T. Modelling anchor text retrieval in book search based on back-of-book index[C]//Proceedings of the SIGIR 2008 Workshop on Focused Retrieval, 2008: 51-58.
- [49] Ratkiewicz J, Menczer F. Text snippets from the DomGraph[C]//Proceedings of the SIGIR 2008 Workshop on Focused Retrieval, 2008: 45-50.
- [50] Bessai-Mechmache F Z, Alimazighi Z. Aggregated search in XML documents[J]. Journal of Emerging Technologies in Web Intelligence, 2012, 4(2): 181-188.
- [51] Jahn N, Lösch M, Horstmann W. Automatic aggregation of faculty publications from personal Web pages[J/OL]. Code4lib, 2010(11) [2016-03-25]. [http://journal.code4lib.org/articles/3765?utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed%3A+c4lj+\(The+Code4Lib+Journal\)](http://journal.code4lib.org/articles/3765?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+c4lj+(The+Code4Lib+Journal)).
- [52] Blake C. Beyond genes, proteins, and abstracts: identifying scientific claims from full-text biomedical articles[J]. Journal of Biomedical Informatics, 2010, 43(2): 173-189.
- [53] Zhang L. Grasping the structure of journal articles: utilizing the functions of information units[J]. Journal of the American Society for Information Science and Technology, 2012, 63(3): 469-480.
- [54] Liu J. Information diversity in Web search[D]. Illinois: Northwestern University, 2009.
- [55] Lin C J, Liu R R. An analysis of multi-focus questions[C]//Proceedings of the SIGIR 2008 Workshop on Focused Retrieval, 2008: 30-36.
- [56] Wan S, Paris C, Krumpholz A. From aggravated to aggregated search: improving utility through coherent organizations of an answer space[C/OL]//Proceedings of the SIGIR 2008 Workshop on Aggregated Search, 2008: 3 [2013-03-25]. [https://www.researchgate.net/profile/Alexander\\_Krumpholz/publication/228520993\\_From\\_aggravated\\_to\\_aggregated\\_search\\_Improving\\_utility\\_through\\_coherent\\_organisations\\_of\\_an\\_answer\\_space/links/0912f50a189a21961f000000.pdf](https://www.researchgate.net/profile/Alexander_Krumpholz/publication/228520993_From_aggravated_to_aggregated_search_Improving_utility_through_coherent_organisations_of_an_answer_space/links/0912f50a189a21961f000000.pdf).
- [57] Sushmita S, Lalmas M. Using digest pages to increase user result space: preliminary designs[C/OL]//Special Interest Group on Information Retrieval 2008 Workshop on Aggregated Search [2016-03-25]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.418.6067&rep=rep1&type=pdf>.
- [58] Ou S, Khoo C S G. Aggregating search results for social science by extracting and organizing research[C/OL]//SIGIR 2008 Workshop on Aggregated Search [2016-03-25]. [http://rgcl.wlv.ac.uk/papers/ou\\_khoo.SIGIR\\_workshop\\_aggregated\\_search.pdf](http://rgcl.wlv.ac.uk/papers/ou_khoo.SIGIR_workshop_aggregated_search.pdf).
- [59] Kaptein R, Marx M. Focused retrieval and result aggregation with political data[J]. Information Retrieval, 2010, 13(5): 412-433.

- [60] Jain R. Storytelling with big multimedia data: keynote talk [C]//ACM. Proceedings of the First International Workshop on Internet-Scale Multimedia Management, 2014: 1.
- [61] 刘植惠. 知识基因理论新进展[J]. 情报科学, 2003, 21(12): 1243-1245. (Liu Zhihui. About the new development of knowledge gene theory[J]. Information Science, 2003, 21(12): 1243-1245.)
- [62] 刘植惠. 知识基因理论研究[J]. 重庆图情研究, 2003(3): 4-10. (Liu Zhihui. About study on the knowledge gene theory[J]. Chongqing Library and Information Science Research, 2003(3): 4-10.)
- [63] 刘植惠. 知识基因理论的由来、基本内容及发展[J]. 情报理论与实践, 1998(2): 8-13. (Liu Zhihui. The origin, content and development of knowledge gene theory [J]. Information Studies: Theory & Application, 1998(2): 8-13.)
- [64] 文庭孝, 侯经川, 龚蛟腾, 等. 中文文本知识元的构建及其现实意义[J]. 中国图书馆学报, 2007, 33(6): 91-95. (Wen Tingxiao, Hou Jingchuan, Gong Jiaoteng, et al. On the construction of Chinese text knowledge elements and its practical significance[J]. Journal of Library Science in China, 2007, 33(6): 91-95.)
- [65] 文庭孝. 知识单元的演变及其评价研究[J]. 图书情报工作, 2007, 51(10): 72-76. (Wen Tingxiao. Research on knowledge unit evolution and evaluation[J]. Library and Information Service, 2007, 51(10): 72-76.)
- [66] 温有奎, 温浩, 徐端颐, 等. 基于创新点的知识元挖掘[J]. 情报学报, 2005, 24(6): 663-668. (Wen Youkui, Wen Hao, Xu Duanyi, et al. Knowledge element mining in knowledge management [J]. Journal of the China Society for Scientific and Technical Information, 2005, 24(6): 663-668.)
- [67] 温有奎. 基于“知识元”的知识组织与检索[J]. 计算机工程与应用, 2005(1): 55-57. (Wen Youkui. The knowledge organization and searches with the knowledge element [J]. Computer Engineering and Applications, 2005(1): 55-57.)
- [68] 温有奎, 温浩, 徐端颐, 等. 基于知识元的文本知识标引[J]. 情报学报, 2006, 25(3): 282-288. (Wen Youkui, Wen Hao, Xu Duanyi, et al. Text knowledge indexing based on knowledge element [J]. Journal of the China Society for Scientific and Technical Information, 2006, 25(3): 282-288.)
- [69] 郑彦宁, 化柏林. 句子级知识抽取在情报学中的应用分析[J]. 情报理论与实践, 2011, 34(12): 1-4. (Zheng Yanning, Hua Bolin. Analysis of the application of the sentence-level knowledge extraction in Information Science [J]. Information Studies: Theory & Application, 2011, 34(12): 1-4.)
- [70] 丁君军, 郑彦宁, 化柏林. 基于规则的学术概念属性抽取[J]. 情报理论与实践, 2011, 34(12): 10-14. (Ding Junjun, Zheng Yanning, Hua Bolin. Extraction of academic concept attribute based on rules [J]. Information Studies: Theory & Application, 2011, 34(12): 10-14.)
- [71] 杨硕, 崔蒙, 赵英凯, 等. 基于知识元的中医药信息知识标引[J]. 中国中医药信息杂志, 2011(8): 24-25. (Yang Shuo, Cui Meng, Zhao Yingkai, et al. Knowledge index about TCM information based on knowledge element [J]. Chinese Journal of Information on Traditional Chinese Medicine, 2011(8): 24-25.)
- [72] 蒋玲. 面向学科的知识元标引关键技术研究[D]. 武汉: 华中师范大学, 2011. (Jiang Ling. Research on key issues in subject-oriented knowledge element indexing [D]. Wuhan: Central China Normal University, 2011.)
- [73] 贾生. 基于知识元的文献知识仓库构建研究[D]. 大连: 大连理工大学, 2012. (Jia Sheng. The research of literature knowledge warehouse construction based on knowledge element [D]. Dalian: Dalian University of Technology, 2012.)
- [74] 付蕾. 知识元标引系统的设计与实现[D]. 武汉: 华中师范大学, 2009. (Fu Lei. Design and implementation of knowledge element indexing system [D]. Wuhan: Central China Normal University, 2009.)
- [75] 葛连兵. 学术文献知识元数据库管理平台的设计与应用[D]. 北京: 清华大学, 2006. (Ge Lianbing. Design

- and application on academic document knowledge element database management platform[D]. Beijing: Tsinghua University, 2006.)
- [76] 姜永常. 基于知识元的知识仓库构建[J]. 图书与情报, 2005(6): 73-74. (Jiang Yongchang. Construction of knowledge warehouse based on knowledge element[J]. Library and Information, 2005(6): 73-74.)
- [77] 陈定权. 论知识链接的建立规则[J]. 图书情报工作, 2010(12): 41-45. (Chen Dingquan. Discussion on linking rules for setting knowledge linking[J]. Library and Information Service, 2010(12): 41-45.)
- [78] 贺德方. 知识链接发展的历史、未来和行动[J]. 现代图书情报技术, 2005(3): 11-15. (He Defang. Knowledge linking: history, future and action[J]. New Technology of Library and Information Service, 2005(3): 11-15.)
- [79] 温有奎, 徐国华. 知识元链接理论[J]. 情报学报, 2003, 22(6): 665-670. (Wen Youkui, Xu Guohua. Knowledge element linking theory[J]. Journal of the China Society for Scientific and Technical Information, 2003, 22(6): 665-670.)
- [80] 姜永常. 基于知识元语义链接的知识网络构建[J]. 情报理论与实践, 2011(5): 50-53. (Jiang Yongchang. Knowledge network construction based on semantic linking of knowledge element[J]. Information Studies: Theory & Application, 2011(5): 50-53.)
- [81] 朱嘉贤, 白伟华, 李吉桂. Web资源的多粒度语义标注及其应用技术研究[J]. 计算机科学, 2011, 38(8): 83-87. (Zhu Jiaxian, Bai Weihua, Li Jijia. Research on the multi-granularity semantic annotation of Web resource and its application technology[J]. Computer Science, 2011, 38(8): 83-87.)
- [82] 陶善菊, 刘清堂, 王凡, 等. 基于知识元的教育技术学科资源库构建[J]. 现代教育技术, 2011(5): 115-120. (Tao Shanju, Liu Qingtang, Wang Fan, et al. Construction of resource base of the education technology based on knowledge element[J]. Modern Educational Technology, 2011(5): 115-120.)
- [83] 丁侃. 基于知识元的中医古籍方剂知识表示研究[D]. 北京: 中国中医科学院, 2012. (Ding Kan. Study on representation based on element knowledge of prescription knowledge in ancient TCM books[D]. Beijing: China Academy of Chinese Medical Sciences, 2012.)
- [84] 温有奎, 焦玉英. Wiki知识元语义图研究[J]. 情报学报, 2009(6): 870-876. (Wen Youkui, Jiao Yuying. Semantic graph of knowledge unit on Wiki platform[J]. Journal of The China Society for Scientific and Technical Information, 2009(6): 870-876.)
- [85] 郭亮. 基于知识单元的语义Wiki知识库[D]. 西安: 西安电子科技大学, 2009. (Guo Liang. A semantic-Wiki knowledge base system based on knowledge elements[D]. Xi'an: Xidian University, 2009.)
- [86] 温有奎, 孙明, 温浩, 等. 基于Web的情报知识元挖掘与语义集成地图[J]. 情报学报, 2008, 27(2): 163-168. (Wen Youkui, Sun Ming, Wen Hao, et al. Information knowledge element mining based on Web and semantic integrating map[J]. Journal of the China Society for Scientific and Technical Information, 2008, 27(2): 163-168.)
- [87] 李建伟. 基于知识元的突发事件情景研究[D]. 大连: 大连理工大学, 2012. (Li Jianwei. Emergency scenario research based on knowledge element[D]. Dalian: Dalian University of Technology, 2012.)
- [88] 钞柯. 基于知识元的突发事件连锁反应模型研究[D]. 大连: 大连理工大学, 2012. (Chao Ke. Research on emergency chain reaction model based on knowledge unit[D]. Dalian: Dalian University of Technology, 2012.)
- [89] 仲秋雁, 郭艳敏, 王宁, 等. 基于知识元的非常规突发事件情景模型研究[J]. 情报科学, 2012(1): 115-120. (Zhong Yanqiu, Guo Yanmin, Wang Ning, et al. Research on unconventional emergency scenario model based on knowledge element[J]. Information Science, 2012(1): 115-120.)

- [90] 黄红雨. 基于知识元的应急案例表示及检索方法研究[D]. 大连:大连理工大学,2013.(Huang Hongyu. Research on method of emergency case representation and retrieval based on knowledge element[D]. Dalian: Dalian University of Technology,2013.)
- [91] 吴悠. 基于知识元的应急决策活动基元模型研究[D].大连:大连理工大学,2012.(Wu You. Research on basic unit model of emergency decision-making activity based on knowledge unit[D]. Dalian: Dalian University of Technology,2012.)
- [92] 刘乃朋. 基于知识元的应急管理模型表示及关联研究[D].大连:大连理工大学,2013.(Liu Naipeng. Research on emergency management model representation and model relevance based on knowledge unit[D]. Dalian: Dalian University of Technology,2013.)
- [93] 董恩超. 基于知识元的非常规突发事件演化模型研究[D].大连:大连理工大学,2012.(Dong Enchao. Study on evolution model of unconventional incident based of knowledge element[D]. Dalian: Dalian University of Technology,2012.)
- [94] 肖文辉. 非常规突发事件知识元获取及知识元网络模型[D].大连:大连理工大学,2013.(Xiao Wenhui. Knowledge unit acquisition and knowledge unit network model of unconventional emergency[D]. Dalian: Dalian University of Technology,2013.)
- [95] 杨德宽. 基于知识元的模型表示与组合调用研究[D].大连:大连理工大学,2013.(Yang Dekuan. Research on representation and integration of model based on knowledge-element[D]. Dalian: Dalian University of Technology,2013.)
- [96] 宋歌. 网络分析方法在引文分析中的整合研究[J].中国图书馆学报,2011,37(94):106-114.(Song Ge. Integration of network analysis methods in citation analysis[J].Journal of Library Science in China,2011,37(94): 106-114.)

曹树金 中山大学资讯管理学院教授,博士生导师。广东 广州 510006

马翠嫦 中山大学图书馆副研究馆员。广东 广州 510275

(收稿日期:2015-12-19)