

“面向众创的科技情报共享服务平台”建设研究

刘 耀 朱礼军 黄 毅

(中国科学技术信息研究所, 北京 100038)

摘要: 针对科技情报(信息)存在的问题, 提出基于国家工程技术图书馆的公益类馆藏资源和特色资源等整合各类数据库及资源, 建设面向众创的科技情报共享服务平台, 以推动实现科技资源综合利用、共享与合作。详细介绍了资源交换与统计、多源信息整合处理、一体化搜索引擎3个核心系统的建设情况, 并分析总结了当前存在的问题与挑战, 旨在为面向众创的科技情报共享服务平台的建设研究提供有益参考。

关键词: 双创; 科技情报; 共享服务; 服务平台; 平台建设

中图分类号: TP391; G35

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2017.04.007

Research on Construction of Scientific and Technical Information Sharing and Service Platform for Mass Innovation

LIU Yao, ZHU Lijun, HUANG Yi

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: “Public entrepreneurship, innovation” is an important measure for our country to implement the innovation-driven development strategy. However, there are lots of issues in the scientific and technical information. In view of this, based on resources of the national engineering and technical library and other related public resources, including patent, paper, and expert database, we put forward the construction of scientific and technical information sharing and service platform for mass innovation, to promote the realization of comprehensive sharing, cooperation and utilization of scientific and technical resources. This paper introduces the construction of the core system of resource exchange and statistics, multi-source information integration and integration search engine, and analyzes and summarizes the existing problems and challenges. It aims to provide a useful reference to the construction of scientific and technical information sharing and service platform for mass innovation.

Keywords: mass innovation, scientific and technical information, sharing service, service platform, platform construction

在国家大力推进“大众创业、万众创新”和推动实施“互联网+”行动的形势下, 科技情报(信息)机构在做好信息搜集、整理、加工、分析以及服务的同时, 应为创新创业者提供产业及

市场咨询、专业技术、互联网技术、科技资金帮扶等信息服务^[1-3]。国家工程技术图书馆的主要任务是根据国家发展需要, 科学、完整地收藏工程技术领域的科技文献信息资源。鉴于此, 笔者

作者简介: 刘耀(1972—), 男, 中国科学技术信息研究所工程中心副主任、研究员, 主要研究方向: 自然语言处理、知识工程; 朱礼军(1974—), 男, 中国科学技术信息研究所工程中心常务副主任、研究员, 主要研究方向: 智能信息处理、知识组织与知识工程、移动问答、语义网等; 黄毅(1986—), 男, 中国科学技术信息研究所硕士研究生, 主要研究方向: 本体、知识工程。

收稿日期: 2017年7月13日。

拟对国家工程技术图书馆的公益类馆藏资源和特色资源等,综合利用多年累积的技术经验,通过整合包括专利库、论文库及人才数据库等各类数据库及相关科技资源,设计建设面向众创的科技情报共享服务平台^[7-11]。

1 平台架构

“面向众创的共享服务平台”采用标准三层架构,分别为应用软件层、平台支撑层、基础设施层(图1)。应用软件层部署门户及通过浏览器访问的各类系统、软件;平台支撑层配置缓存服务、集群、中间件、海量计算、身份验证,数据库服务等;基础设施层配置了数据存储环境,服务器、主机、虚拟化、网络设备、负载均衡等。该平台按照“技术工具化、工具业务化、业务自动化”的设计思想,将异构分布的多源信息进行有效整合,通过统一的、标准的信息接口,为众创活动的主体提供包括数据交换、资源检索、分析统计、综合评价以及自动报告等在内的科技情报服务。

2 资源交换与统计系统

“面向众创的共享服务平台”的资源体系主要由基础文献、辅助文献和二次开发文献三大类

构成。其中,基础文献包括国家工程技术图书馆的科技期刊、论文、专利、标准和科技报告等;辅助文献包括科技成果、科技政策、行业动态等;二次开发的文献包括专家库、项目库、技术专题库等。这些异构资源存在多种形式、多种格式,颗粒度也不一样,将这些资源进行统一的处理,形成标准的XML文档,是科技信息资源共享与应用的前提。

资源交换与统计主要是通过互联网、数据库及本地等多种渠道获取资源,再对不同来源、不同层次、不同结构的资源进行识别与选择,依据构建的Schema标准,对资源的形式进行语义化处理,将解析后的XML资源加入数据交换池,以统一的接口对外提供数据交换服务,并通过统计分析用户日志,实现用户行为分析及资源计费管理。

2.1 标准化体系

信息标准化是技术、产品、应用系统研发以及信息技术服务的前提。遵循标准化体系设计与建设的基本原则,保障遵循及引用标准规范的应用系统和技术系统安全、可靠、稳定运行。科技情报共享服务平台支持 Schema 模板上传、模板解析、模板列表显示、模板预览、模板修改、模板导出、模板删除等功能。其中,模板修改主要

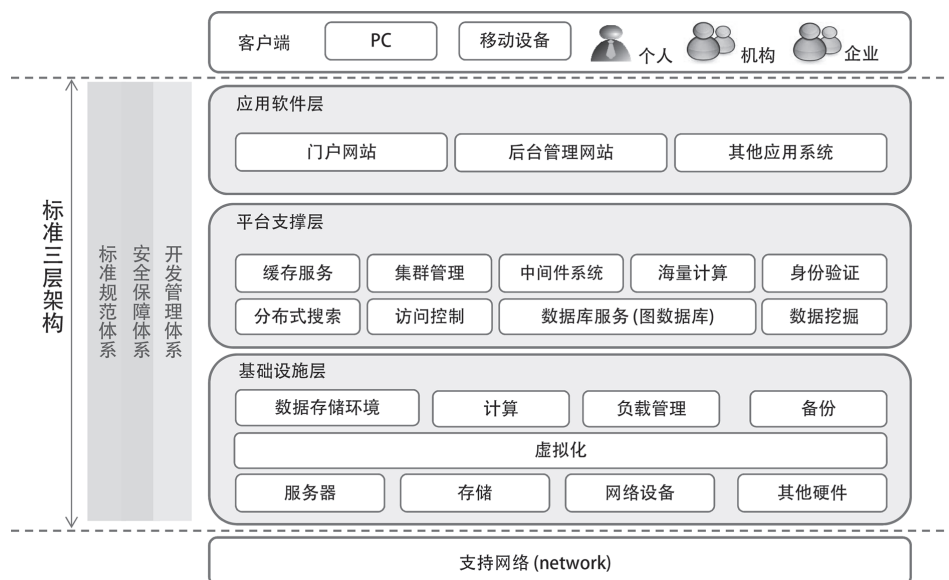


图1 平台技术架构图

是对 schema 文件的编辑,包括增加、修改、删除元素及相关属性等。该平台以现有国内外相关标准体系为基础,采取图书、论文、工具书、专利等各种资源类型的 Schema 标准,建立动态更新的标准体系服务平台。软件研发人员、集成人员、各类用户及政策决策者等通过该标准体系服务平台能够对数据进行按需定制化处理。

2.2 数据交换

采用应用广泛且普及的 XML 来解决不同系统之间的异构性,通过设置全局对象 Schema 对图书、专利、论文、期刊等文献资源的节点和属性进行解析,提供 Web Service 的数据检索接口和数据更新接口。用户可以通过调用接口对资源实现查询以及更新,解决资源的模式异构、资源的语义异构以及不同格式所带来的代码冗余和性能损失问题,并通过定制资源的输出结果和格式完成相关数据的交换。

数据交换主要包括格式解析转化和数据内容交换两个部分。通过解析 Schema、配置 solr 字段(普通域、动态域、复制域)对资源解析、索引进行定制。科技情报共享服务平台具有添加、修改、删除普通域、动态域的功能。根据 Schema 解析数据资源,对图书、专利、论文等多种形式等资源的元数据、标签、文本内容的自适应与精准解析与索引,实现对资源形式语义处理过程,

并进行统一存储,提供 Web Service 数据查询、导出等服务。

2.3 统计分析

通过对平台相关数据的统计分析可以实现用户行为的分析和资源的计费管理(图2)。通过计算机建模和数据挖掘的方法,分析用户行为的内在动机,建立用户参与行为模型,从而有效揭示宏观舆论与微观个体行为的联系,有助于把握行业动态,发现潜在的热点话题,为创新活动主体决策提供技术手段。用户行为分析主要包括初步过滤用户行为事件、定义用户行为、分析与判定用户行为、匹配用户行为、更新行为存储数据库、分析与判定行为等。通过对用户操作日志进行统计、数据挖掘和分析,实现个人、群组用户管理,加强操作日志管理。依据用户日志统计分析,还可以实现数据的可视化展示,包括系统访问量、平均停留时间、检索频次、检索关键词、资源下载次数、平均活动次数、每次单访问最大活动量的信息的可视化展示,实现趋势分析、访问日志分析、忠诚度分析、访问时间分析等多种功能。科技情报共享服务平台还可以提供资源计费管理,并依据资源使用方式、资源来源、资源使用量进行结算。通过多角度的资源使用统计和分析,能够深入了解创新主体对平台资源的需求和使用情况,为资源进行价值评估提供辅助依

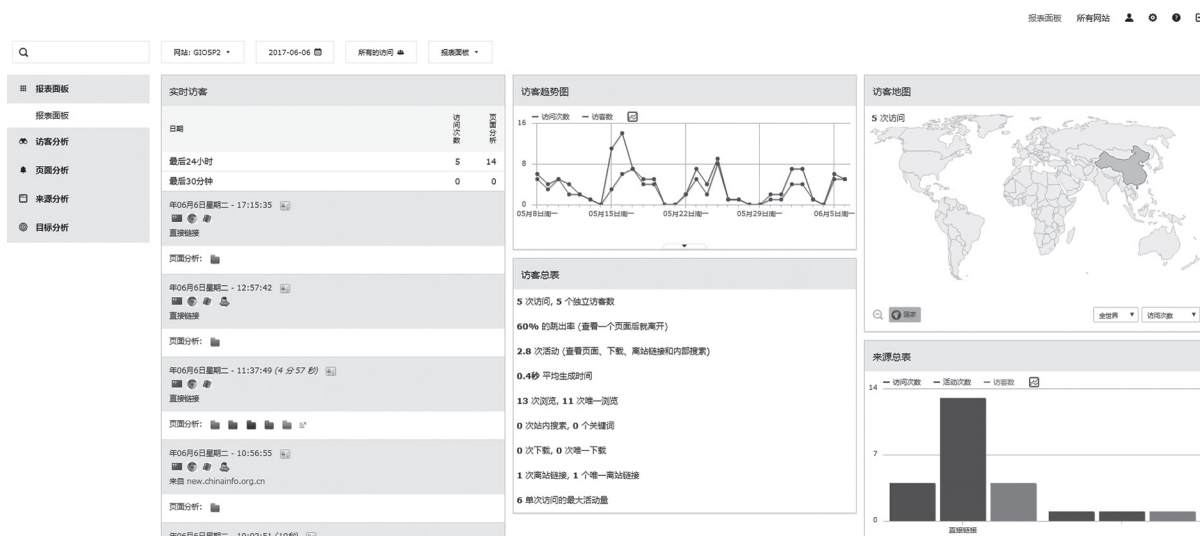


图2 用户行为分析

据,有助于实现对资源的高效利用和管理。

3 多源信息整合处理系统

该系统从资源交换与统计系统中获取数据资源,对文本、表格、图像等不同模态的资源整合以及内容进行语义化处理,对数据进行知识抽取、内容重组等二次加工形成新的资源,再将解析处理后的数据资源存储到数据交换与统计系统中。通过与科技情报共享服务平台的数据交换与统计系统的融合与集成,为一体化搜索引擎提供接口,为资源的交换与复用奠定基础。

3.1 资源整合

通过Web Service数据接口从资源交换与统计系统中获取数据资源(图3),实现资源的整合。

3.2 资源处理

根据创新主体资源组织结构化的需求,在通用自然语言处理技术的基础上,通过面向资源处理的定制化的管线技术,实现处理工具的自由组合和定制,从而解决资源处理需求多元化、多样化的问题。通过接口获取数据资源,并对其进行分段、分句、分词、句法分析、语义标注等内容语义化的处理,实现多源信息的有机融合,使其具有较强的柔性、条理性、系统性。同时建立资源索引,通过对已有资源内容进行深度挖掘、知识抽取与内容重组,生成新的资源库内容,如专家库、项目库、技术库等,提升资源的价值。

3.3 资源管理

通过对词典、规则模板、专家库、项目库等资源的管理,以及对相关数据进行增、删、改、查等操作,实现对科技情报共享服务平台的资源

管理。在资源管理过程中,建立同义词词典,以提高数据的查全率;建立停用词词典,以提高检索效率;对规则模板库的管理,以维护和更新知识抽取规则、解析规则、内容重组规则;对专家库、项目库、技术库进行管理,及时维护和更新相关资源。通过调用接口,将解析处理后的资源存储到资源交换与统计系统中。

4 一体化搜索引擎系统

通过与多源信息整合处理系统和资源交换与统计系统的融合与集成,按照不同颗粒度(篇、章节、段落、句子级别)对资源进行标注与索引,建立多维语义索引体系,构建一体化智能检索引擎,支持多种检索条件的自由组合及分析,并针对众创的需求提供找文献、找技术、找专家、找项目等精准检索服务。

4.1 找文献服务

找文献服务主要实现了对科技文献资源的快速、精准检索,提供了语义检索、高级检索、报告生成等服务(图4)。进入“找文献”检索首页后,在检索框中输入相关检索词,点击检索按钮进入检索结果列表页面,可以按资源类别、语义分类进行筛选,可以按时间、热度等进行排序,也可以按时间、区域、技术分类等生成整体概况统计分析趋势图,从宏观层面了解行业、技术的整体概况还可以进行专利相似度对比、专利文本分析、知识图谱分析等,从中观和微观层面了解行业、技术的情况,甚至可以围绕主题生成分析报告,可以涵盖时间趋势分析、区域分析、技术分类分析、专利排行等内容,并辅助用户在此基础上快速形成行业、技术分析报告。

4.2 找技术服务

找技术服务主要实现对技术的快速查找与分析,从宏观层面了解技术的发展阶段和趋势,从中观和微观层面了解技术的热点与空白点。在找技术主页,可以提供技术曲线分析、功效矩阵分析、聚类分析、技术百科等服务。

通过一套全面的包含技术生命周期发现、技术成熟度预测等模型方法和体系,综合分析会议

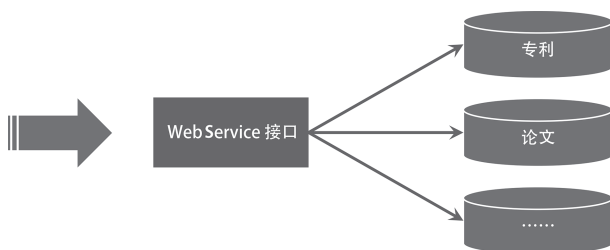


图3 资源获取

论文、期刊论文及专利申请的时间顺序及数量等特征，并针对任意检索词自动生成相关主题的生命周期曲线。在找技术检索主页，点击技术曲线，进入技术曲线页面，输入检索词，点击搜索按钮，即可生成与检索词相关的技术曲线（图5）。

以每一篇科技文献文本中的术语作为技术词、功效词，并结合领域本体、语义标注索引技术较全面、准确地确定技术、功效的特征，统计

其特征及对应包含文献数量，形成动态技术功效矩阵，实现点击功效矩阵中横轴与纵轴上的信息点，动态变化生成新的功效矩阵，以提供更加精细的技术空白区和技术研发热点的分析。在找技术检索主页中，点击功效矩阵，进入功效矩阵页面，可以提供生成与某一主题相关的关键技术、功效所对应的专利多维矩阵图的功能（图6）。

利用文本挖掘技术，结合专利文本的特征，



图4 语义检索结果

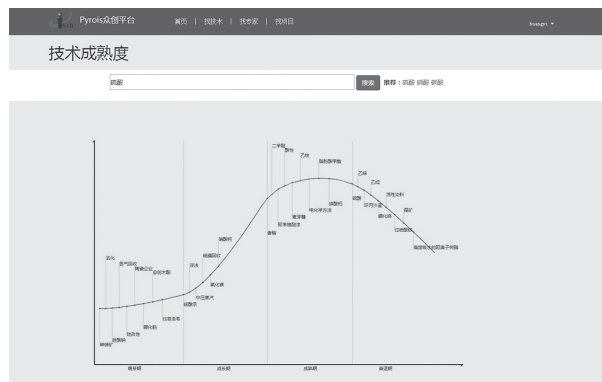


图5 技术曲线页面

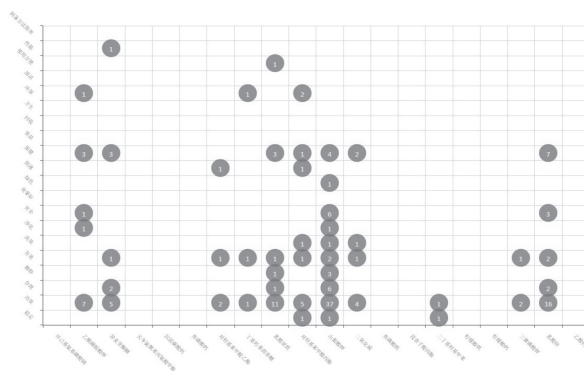


图6 技术功效矩阵页面

采用基于多维标度法的专利文本可视化聚类方法,通过分析专利之间的相似度及关联程度,采用等高线刻画专利数量,形成地形图,对当前专利技术领域的应用主题分布状态进行可视化展示。在找技术检索主页,点击聚类分析进入聚类分析页面,输入检索词点击分析按钮,可以生成聚类分析图(图7)。

依据构建的深度融合的多源信息知识网络,可以自动生成技术百科。在找技术检索主页,输入检索词,点击技术百科,即可进入技术百科页面(图8)。

4.3 找专家服务

找专家服务通过建立的专家评价、认证与沟通体系，与创新主体建立连接当创新主体在开展创新活动或遇到技术瓶颈时，能够通过科技情报共享服务平台查找技术、行业专家为他们提供技术评估、咨询服务。在找专家主页，提供了检索研究人员、技术人员、设计人员的入口。在找专家主页，如输入检索词，点击检索，即可进入专

家检索结果页面(图9)。然后点击专家姓名,即可进入专家详情页面(图10)。在找专家主页,如点击研究人员可以进入研究人员检索主页,还可以按检索条件进行检索。

4.4 找项目服务

找项目服务主要对各类项目提供查询服务。在找项目主页, 提供了检索科研项目、投资项目入口。当输入如人工智能检索词后, 点击检索, 即可进入科研项目检索结果页面 (图 11)。

在找项目主页，当点击科研项目后，进入科研项目检索主页，然后输入检索条件，如：国家自然科学基金项目，点击检索，即可进入科研项目检索结果页面。再在科研项目检索结果页面中点击项目名称，即可进入该项目详情页面（图 12）。

5 应用效果

“面向众创的科技情报共享服务平台”通过整合各类文献及事实数据等资源，研发了资源交

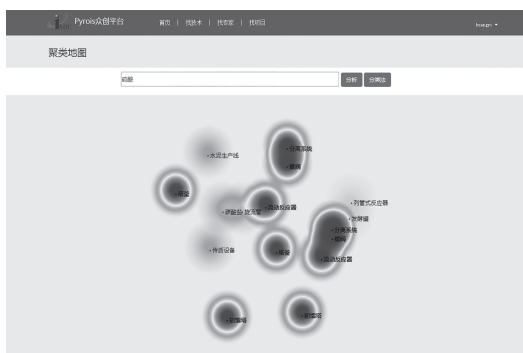


图 7 聚类分析页面

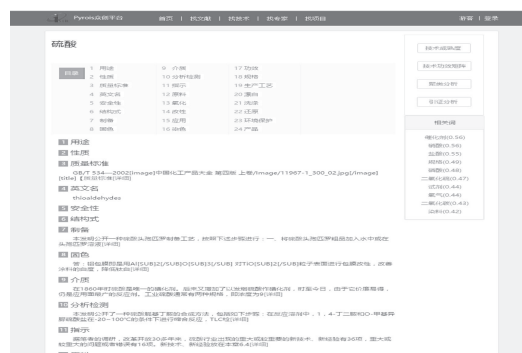


图 8 技术百科页面



图9 专家检索结果页面



图 10 专家详情页面

#	项目编号	项目名称	所属领域	立项年份	项目状态	经费/万元
1	60525207	微纳科学部 微纳印刷、人工智能	2004	进行中	南京工大	100.0
2	60726203	微纳科学部 微纳印刷、人工智能	2008	进行中	中国科学院上海硅基研究所	40.0
3	60426202	微纳科学部 微纳印刷、人工智能	2005	结束	中国科学院技术大学	40.0
4	60525202	微纳科学部 微纳印刷、人工智能	2006	结束	浙江大学	100.0
5	60625203	微纳科学部 微纳印刷、人工智能	2000	进行中	北京大学	80.0
6	61222110	微纳科学部 人工智能制造及应用	2013	进行中	浙江大学	100.0
7	6041020246	微纳科学部 微纳印刷、人工智能制造及应用	2004	结束	中国科学院材料研究所	0.0
8	60116119	微纳科学部 2007年人工智能制造及应用	2001	结束	北京理工大学	0.0
9	02826025	微纳科学部 人工智能制造及应用	2002-07-01	进行中	山东大学	0.0
10	20477010	微纳科学部 微纳印刷、人工智能制造及应用	2005	结束	华东理工大学	22.0
11	04820046	微纳科学部 微纳印刷、人工智能制造及应用	2004-05-05	进行中	清华大学	0.0
12	10C20014	微纳科学部 微纳印刷、人工智能制造及应用	2010-06-17	进行中	山东大学	0.0
13	61473212	微纳科学部 微纳印刷、人工智能制造及应用	2014	进行中	清华大学	67.0
14	6021020206	微纳科学部 微纳印刷、人工智能制造及应用	2002	结束	中国科学院材料研究所	0.0
15	60010300304	微纳科学部 微纳印刷、人工智能制造及应用	2006	进行中	清华大学	3.0

图 11 科研项目检索结果页面

项目信息	项目详情
项目名称	基于DNA自动机过程的双链DNA序列测定的新方法
项目编号	19FGL001
专业领域	国家自然科学基金项目
承担单位	天津科技大学
项目类别	生命科学部
项目负责人	张洪刚
项目起止时间	2006
项目经费	2006
项目状态	进行中
项目地址	天津科技大学
项目人员	张洪刚
项目成果	张洪刚

图 12 项目详情页面

换与统计系统、多源信息整合处理系统以及一体化搜索引擎，构建了一个面向众创的科技情报资源库。该数据库的资源类型包括专利、论文、图书、科技报告、机构、成果、项目、人才等，资源总量约 1.5 亿条。目前，该平台通过植入的方式，已在甘肃省科技厅、厦门产业研究院等单位开展了应用示范工作，服务对象包括 815 个科研店铺、33 个众创空间、10 余家企业、4 家行业协会等，能够面向创新创业主体在创新链上的不同环节、不同角色、创新过程中遇到的问题，有针对性的提供科技情报、知识资源的导航、查询与分析服务，加强了创新创业主体与企业、产业链、行业协会的之间的交流、合作与共享，有效地改善了创新创业环境，提升了创新能力与效率。

6 结语

本文首先设计了国家工程技术图书馆面向众创的科技情报共享服务平台的架构，然后探讨资源交换与统计、多源信息整合处理、一体化搜索引擎 3 个核心系统的建设情况及取得的成效，最后面对新形势下的挑战提出相应的对策和建议。

(1) 国家工程技术图书馆共享服务平台将专利、论文、图书、科技报告等创新资源进行深度融合，为用户提供有针对性的综合查询，并提供深度分析及报告工具支持，以帮助创新主体掌握相关科研、专利、产品等方面的发展现状与趋

势，为创新活动提供基础决策数据支持。该平台依据众创概念模型设计，将优秀的人才、项目、团队汇聚在一起，建立起创新主体、技术、项目、专家之间的联系，形成一个专业的科技情报集散枢纽。

(2) 国家工程技术图书馆共享服务平台提供的主要是信息资源服务，应充分利用云计算、物联网、移动互联网、大数据等新兴技术，将通用的技术、资源、系统以服务模块的方式应用于平台建设中，积极提供线上流程定制服务。

(3) 国家工程技术图书馆共享服务平台还应围绕信息技术、生命健康、智能制造等战略性新兴产业，深度融合包括产业、市场、企业等在内的多源信息，提供行业发展态势的分析与预测、企业竞争情报、技术先进性评价、创业团队评价等服务。

(4) 目前国家工程技术图书馆共享服务平台的影响范围较小，应通过媒体及各种渠道，组织各种活动对平台进行推广宣传，加大推广宣传力度，提高平台在双创活动主体中的知名度，以真正发挥共享服务平台应有的作用。

参考文献

- [1] 汪保君.论区域创新的科技文献信息资源整合和共享平台建设[J].图书馆,2008(2):69-72.
- [2] 王洪波,黄倩,张鹤.基于服务创新视角的科技文献共享平台建设研究:以武汉科技信息共享服务平台

- 为例[J]. 中国科技资源导刊, 2017, 49(1): 38-41.DOI: 10.3772/j.issn.1674-1544.2017.01.006.
- [3] ZHAO Jing, CHENG Yonghong, HUANG Yi. Research and development of scientific and technical information sharing and service platform for mass innovation[J]. ICIC Express Letters, 2017, 8(1): 85-91.
- [4] 肖瑞兰. 河南科技文献信息资源共享服务平台建设研究[J]. 图书与情报, 2009(5): 144-146.
- [5] 谭钢, 梁宏霞. 省、自治区级科技文献共享平台区域特色建设的探索: 以广西科技文献共享与服务平台为例[J]. 图书馆界, 2010(6): 47-49.
- [6] 屈宝强, 吴家喜, 赵伟, 等. 地方科技文献共享平台数据库建设情况分析[J]. 情报杂志, 2011, 30(8): 71-75, 70.
- [7] LIU Yao, CHEN Xuefei, SUI Zhifang, et al. Research on semantic method of library resources' organizing[J]. ICIC Express Letters, 2011, 5(4): 1011-1017.
- [8] LIU Yao, SHI Haiqing, ZHENG Deju. Study on semantic annotation for professional literature[J]. ICIC Express Letters. Part B, Applications, 2014, 5(5): 1383-1389.
- [9] LIU Yao, HUANG Yi, WANG Yan. Research on the key technologies of pyrios knowledge service platform[J]. ICIC express letters. Part B, Applications: an international journal of research and surveys, 2015, 6(5): 1323-1328.
- [10] LIU Yao, ZHANG Ziyuan, HUANG Yi. Research and development of semantic annotation platform for scientific literature[J]. ICIC Express Letters, 2016, 10(7): 1787-1794.
- [11] LIU Yao, LI Yanping, HUANG Yi. Research on semantic and syntactic analysis of patent literature[J]. ICIC Express Letters, 2016, 10(2): 471-477.

(上接第30页)

再使用SWOT分析、定标比超分析、财务报表分析、战争游戏等竞争情报分析方法对竞争环境、竞争对手进行分析, 最终帮助投资人做出合理的决策。

(3) 在“双创”推动下市场竞争更加激烈, 投资者在自身已有信息需求的基础上用情报驱动每一个节点的决策制定, 最终做出由综合情报支持的战略规划, 做出可行的预测性分析。

然而, 本文仅仅是运用多案例研究法对投资者的信息需求进行研究, 还存在很多的不足。比如文中只是进行数据的收集和对比分析, 如果能够将创业者的创业项目信息与投资人在投资之后各个阶段的信息需求进行统计分析, 则可以更加深入准确地分析投资人的信息需求, 结果也将会更具代表性。

参考文献

- [1] 改革创新促进了我国就业持续扩大[EB/OL].[2017-03-29].http://www.gov.cn/xinwen/2017-03/29/content_5181615.htm.
- [2] 辜胜阻, 何峥. 推动“双创”可持续发展[J]. 中国人才, 2016(13): 38-38.
- [3] 洪建玲. 关于企业项目投资决策的竞争情报支持研究[J]. 管理科学文摘, 2008(3): 99-100, 130.
- [4] 江硕远. 基于制度理论的中国天使投资人评价标准研究[D]. 合肥: 合肥工业大学, 2013.
- [5] 雷文武, 左臣真, 苏志敏, 等. 竞争情报在风险投资决策中的应用研究[J]. 商场现代化, 2011(15): 64-65.
- [6] 彭靖里, 宋林清, 郭平, 等. 论投资项目可行性研究中的竞争情报分析[J]. 情报理论与实践, 2003, 26(6): 523-527.
- [7] 李月琳, 胡玲玲. 投资者信息行为分析: 信息源的选择与利用研究[J]. 情报资料工作, 2012(4): 90-97.
- [8] 赵一鸣, 黎苑楚. 面向创业投资的信息集成与服务系统研究[J]. 科技进步与对策, 2012, 29(24): 32-36.
- [9] 潘琰, 辛清泉. 解读企业信息需求: 基于机构投资者的信息需求探索[J]. 会计研究, 2004(12): 14-22.
- [10] 百度百科: 行业分类[EB/OL].[2017-03-12].<http://baike.baidu.com/>.