



中国工程科技知识中心项目

汉语科技词系统

(黑色金属卷)



CHINESE SCIENTIFIC & TECHNICAL
VOCABULARY SYSTEM

FERROUS METALS

- 钢铁研究总院
- 中国科学技术信息研究所 编著
- 中国工程院



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

中国工程科技知识中心建设项目

汉语科技词系统

（黑色金属卷）

钢 铁 研 究 总 院

中国科学技术信息研究所 编著

中 国 工 程 院



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

· 北 京 ·

图书在版编目 (CIP) 数据

汉语科技词系统. 黑色金属卷 / 钢铁研究总院, 中国科学技术信息研究所, 中国工程院编著. —北京: 科学技术文献出版社, 2017. 4

ISBN 978-7-5189-1823-2

I. ①汉… II. ①钢… ②中… ③中… III. ①科学名词—检索系统 ②黑色金属—名词术语—检索系统 IV. ① G254.92

中国版本图书馆 CIP 数据核字 (2016) 第 198006 号



二维码一扫
相关内容尽知晓

汉语科技词系统 (黑色金属卷)

策划编辑: 周国臻 责任编辑: 张 丹 责任校对: 赵 媛 责任出版: 张志平

出 版 者 科学技术文献出版社
地 址 北京市复兴路15号 邮编 100038
编 务 部 (010) 58882938, 58882087 (传真)
发 行 部 (010) 58882868, 58882874 (传真)
邮 购 部 (010) 58882873
官 方 网 址 www.stdp.com.cn
发 行 者 科学技术文献出版社发行 全国各地新华书店经销
印 刷 者 北京教图印刷有限公司
版 次 2017年4月第1版 2017年4月第1次印刷
开 本 787×1092 1/16
字 数 983千
印 张 37.5
书 号 ISBN 978-7-5189-1823-2
定 价 178.00元



版权所有 违法必究

购买本社图书, 凡字迹不清、缺页、倒页、脱页者, 本社发行部负责调换

目 录

第一部分 词系统背景知识	1
1 汉语科技词系统的理论进展	3
1.1 汉语科技词系统的适应性	3
1.2 关系空间和属性空间	3
1.3 面向知识服务应用的词系统建设	4
2 汉语科技词系统协同构建平台进展	6
2.1 集成的新应用	6
2.2 平台功能方面的完善	8
3 汉语科技词系统应用探索	12
3.1 新材料领域	12
3.2 医药卫生领域	13
4 知识结构设计(以金属材料为例)	16
4.1 设计理念	16
4.2 金属材料词系统属性类型说明	16
4.3 金属材料词系统关系类型说明	17
第二部分 黑色金属领域概述	21
1 发达国家现状及趋势	23
2 国内现状与成就	26
2.1 综合情况	26
2.2 行业技术发展	27
2.3 行业技术创新体系建设	28
2.4 产业链建设	30
2.5 发展趋势	31
2.6 现存主要问题	32
2.7 对策及措施	33

第三部分 黑色金属领域汉语科技词系统实例	35
1 格式说明	37
2 实例正文	40
第四部分 附 录	575
A 实例词条音序	577
B 实例词条笔画索引	583

第一部分

词系统背景知识

1 汉语科技词系统的理论进展

汉语科技词系统在经过新能源汽车、新一代工业生物技术、智能材料与结构技术、重大自然灾害监测与防御、新能源等几个领域的系统建设之后,已经基本确定面向情报分析的领域知识组织系统构建的理论方法。自2014年参与中国工程科技知识中心(简称“知识中心”)项目以来,尤其是新材料和皮肤病问答专业领域词系统的建设及对材料测试分析本体的尝试,使得团队对词系统本身的认识及其在知识服务中的应用取得了新的进展,这些进展有助于黑色金属和有色金属两个领域词系统的建设。

1.1 汉语科技词系统的适应性

汉语科技词系统在适应性方面的调整主要是对知识中心的支持。汉语科技词系统设计之初,是为了支撑情报分析服务。但是,词系统也可以支撑知识中心的知识组织系统建设相关工作。知识组织系统建设是知识中心建设和服务过程中的一项重要任务。通过工程及相关领域以术语词汇为核心的知识组织系统建设,汇聚领域内规范化的术语词汇,是高效深入的信息服务和知识服务系统建设中的重要环节。预期通过知识组织系统的建设能够明确知识框架,优化数据资源配置,提高知识加工效率和一致性,增强知识导航和知识服务的效果。通过知识组织系统建设能够更好地组织知识中心的数据资源,为国家工程科技领域重大决策、重大工程科技活动、企业创新与人才培养提供知识服务。

知识中心的协建单位涉及多个行业、多个专业领域,既包括中国农业科学院农业信息研究所、中国医学科学院医学信息研究所、工业和信息化部电子科学技术情报研究所、中国地质图书馆等传统图书情报研究单位,也有如钢铁研究总院、中国环境科学研究院和北京低碳清洁能源研究所等领域专业研究单位。同时,各单位的基础不一,有的长期从事词表、本体等知识组织系统建设,有的过去没有相关建设经验。在这样的现状下,对所有的分中心做统一的要求,建设划一标准的知识组织系统是不现实的。

而汉语科技词系统相关的理论方法和工具平台对构建不同层次的知识组织系统均有支撑。除建设标准的融合叙词表、词典和本体的领域科技词系统外,由于关系、属性的可定义特点使得只需简单配置就可以支持叙词表建设,而词系统本身对属性的支持,可以支撑本体的建设,同时通过相关研究使得词系统向本体转换的工具也得以开发。因此汉语科技词系统可以支撑知识中心从词表到本体的建设,同时由于数据结构的一致性,便于从整体上进行集成整合。

1.2 关系空间和属性空间

汉语科技词系统建设中非常核心的一项工作是关系空间和属性空间的设计,一个好的领域词系统的关系空间和属性空间的建设,是构建词系统的质量、效率及在知识应用中发挥作用的重要保障。

在汉语科技词系统的发展过程中,为了保持体例的一致性及从集成的角度出发,提出了 20 种扩展关系、16 种属性作为推荐的基本类型,并且指出可能存在领域特性,这一思想指导了后续若干领域词系统的建设。但是随着领域词系统的建设向纵深发展,越来越多领域之间的差异性体现得更为明显,为此团队适时调整了原有的思路,提出在领域词系统建设上以领域特色为主、以通用关系为辅的建设思路,以适应不同领域的特点。团队通过调研发现即使在工程科技领域不同学部、学部内的不同学科关注的重点也不尽相同。

关于词系统关系空间和属性空间的大小,也有重新认识的过程,在新能源汽车领域词系统的建设过程中关系包括了 15 个一级类型、78 个二级类型,属性中包括了 10 个一级类型、45 个二级类型。在之后的过程不断扩充,如在黑色金属领域词系统,在建设过程中关系扩展到 109 种,属性类型扩展到 93 种。这些扩展相比只有用、代、属、分、参、族等简单的词和概念关系及一些简单属性,知识在丰富程度上得到了极大的提升,为知识服务提供了较好的基础。但是在建设过程中,知识工程师也反映在近百种关系和属性中进行选择是一件复杂的工作,很多情况下难以找到最合适的,且也难以区分部分属性和关系类型的细微差别。另外,从统计上考虑,存在较大的不平衡性,所以适当减少关系和属性类型有利于减少加工难度。因此,如何选择关系和属性,需要结合应用。

1.3 面向知识服务应用的词系统建设

针对不同的应用,词系统建设应该做针对性分析处理,各类应用包括但不限于如下几类。

1.3.1 数据资源组织和检索

利用建好的汉语科技词系统对数据资源进行标引加工,并基于标引加工的结果提供基本的检索服务。

知识组织服务主要是知识资源的主题标引和多领域标引。主题标引以特定的专业汉语科技词系统为来源,精选代表性词汇加以标注,可以进一步以标注词汇代替资源进行简化计算。多领域标引则以多个汉语科技词系统标引资源中全部有关词汇,并以计算机可理解的形式加以描述,便于进一步利用。

知识检索主要利用汉语科技词系统对用户的检索需求进行明确化交互、对检索结果进行扩检和缩检,对检索结果进行分类展示或聚类展示等。

1.3.2 知识导航

知识导航是利用已有汉语科技词系统的分类体系或高位概念对知识服务提供的科技文献、科学数据等资源和百科、图片等片段化条目知识进行组织,并在知识服务中按照分类体系进行逐层展开的树形关联导引,帮助用户逐步定位找到所需资源和知识。知识导航是重要的交互手段之一,适用于对专业了解较为深入的用户,能够与检索功能有效互补,是知识中心专业分中心的基本建设要求之一。

1.3.3 其他知识服务

知识服务是在对信息资源提炼、加工、集成的基础上,为用户提供的有针对性的、解决具体问题的知识产品或服务的活动。人类从噪声中分拣出数据,转化为信息,升级为知识,升华为智慧,知识服务是信息服务的高级版本,是面向增值和创新的服务。知识服务种类众多,以下仅列举部分。

(1) 针对知识问答提供知识的精准匹配

将重要问答知识与查询词条一一对应,针对查询给出标准的问答结果。

(2) 针对深度搜索提供检索结果的筛选和重新组织

在通用的检索策略得到的检索结果之上,利用已有的汉语科技词系统对查询结果进行再次甄别,识别出不符合检索领域的结果,使得检索结果更加精准,可以进一步用于情报分析。

(3) 针对知识地图提供领域知识概貌

利用汉语科技词系统的主要层次结构,提供对领域知识概览的支持;利用网状结构,提供对领域知识脉络展示支持。

(4) 针对科技评价提供评价对象,界定评价范围

提供主要体现工程科技有关技术的术语作为评价对象,便于确定科技评价中趋势分析等处理的对象范围,做好科技评价工作。

(5) 针对热度发现提供概念关联支撑

提供词汇之间的关联,使得表现不同的词汇之间能够通过概念层面建立联系,可以进一步提高热度分析的准确度。

(6) 针对文本聚类分析,提供聚类特征和降维工具

文本聚类分析中,经常以词汇作为基本特征,但是常常面临多维度问题,而利用汉语科技词系统之间的关联,可以有效降维。

(7) 机器翻译和跨语言检索

对不同语种的资源进行机器翻译,便于对资源进行初选,有助于提供统一检索服务。

2 汉语科技词系统协同构建平台进展

汉语科技词系统协同构建平台经过近些年的发展和实际应用,吸收了一些最新的、成熟的研究成果,采纳了领域用户和行业专家的建议,集成了可视化展示、关系推荐、选词等新的应用,在分领域管理、按用户分配和交互操作便利性等方面对功能进行了完善,目前已经具有较高的成熟性、易用性。

2.1 集成的新应用

最新的汉语科技词系统协同构建平台集成了“汉语科技词系统建设与应用工程”“本体发布与服务平台”和“汉语科技词系统语料库平台”等项目的研究成果,完成了一些新应用的集成工作。在构建领域词表功能上,主要集成了“汉语科技词系统构建与应用工程”的词条的加工流程、词表的管理、用户管理等;在系统的前台浏览检索上,主要集成了“本体发布与服务平台”的关系可视化功能,用户在页面中可以直观地查看词条的关系和属性信息;在后台加工中,主要集成了语料库计算功能,在词表加工人员针对领域词表加工时,可以通过导入领域语料,在计算后获取领域词条,并且导入到词表中;在增加词条关系的时候,可以通过语料库关键词共现功能,获取与当前词条的共现词条,然后建立关系,在增加词条关系时,系统会提示计算后的关键词共现,大幅度提高了加工人员的加工效率(如图 1-1 所示)。

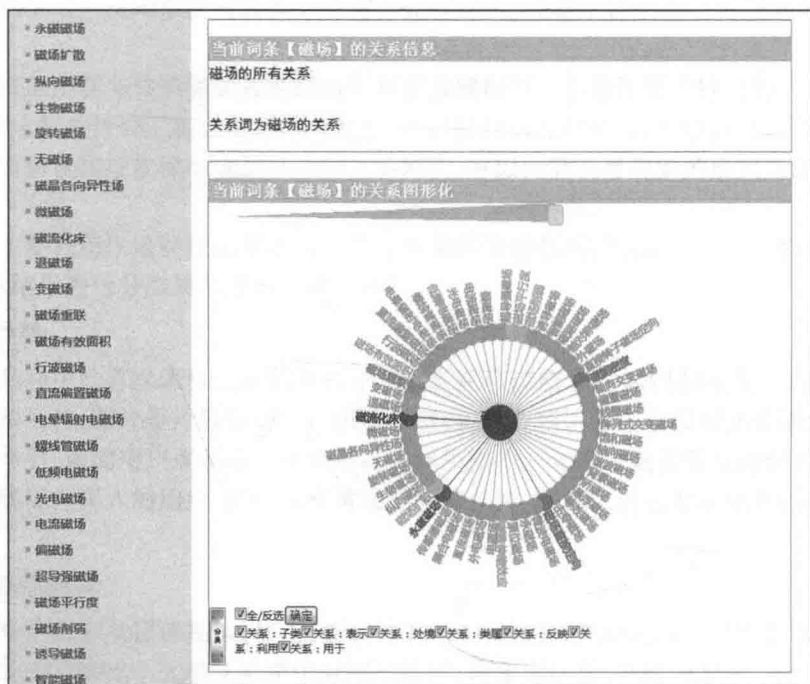


图 1-1 关系可视化功能

其中的语料计算应用主要是实现词语推荐和词条关系推荐功能。用户导入一批语料,并计算语料中关键词的词频,然后导出词频计算结果。用户将结果拿给专家检查后,由专家将检查后,觉得是核心词的词语保留,将修改过的词频结果重新导入到系统中,并计算核心词在语料中的共现情况。在用户增加词条的关系时,将与该词条共现次数较多的词语推荐给用户。语料计算功能,包括“语料管理”(如图 1-2 所示)“数据计算”(如图 1-3 所示)“关键词共现统计”和“关键词词频统计”(如图 1-4 所示)功能,并提供计算结果供用户做选词参考。



图 1-2 语料管理功能

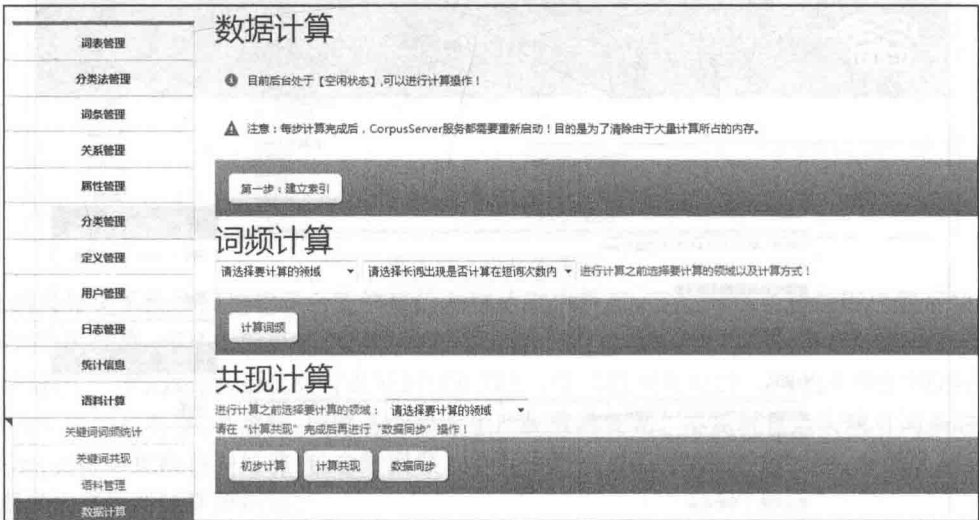


图 1-3 语料计算功能

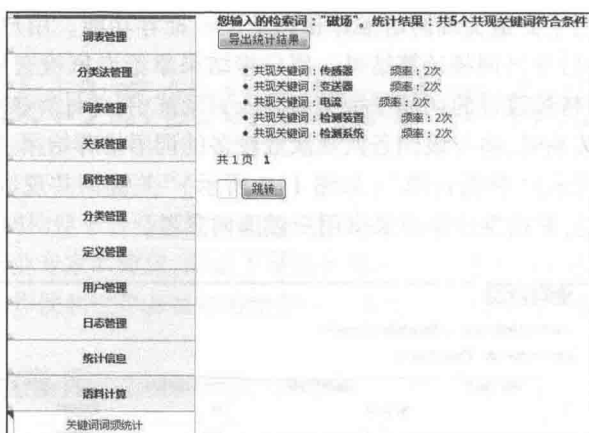


图 1-4 关键词共现统计功能

2.2 平台功能方面的完善

汉语科技词系统协同构建平台在功能上进行了丰富和完善,在分领域管理、交互操作和用户管理分配等方面有了较大的改进。

在分领域管理方面,平台可以同时管理多个领域的词表(如图 1-5 所示)。多个领域的知识工程师和专家可以在一个平台上同时完成不同领域词表的加工、审核和发布等工作。用户也可以对同一个词语在多个领域的词表内进行检索。

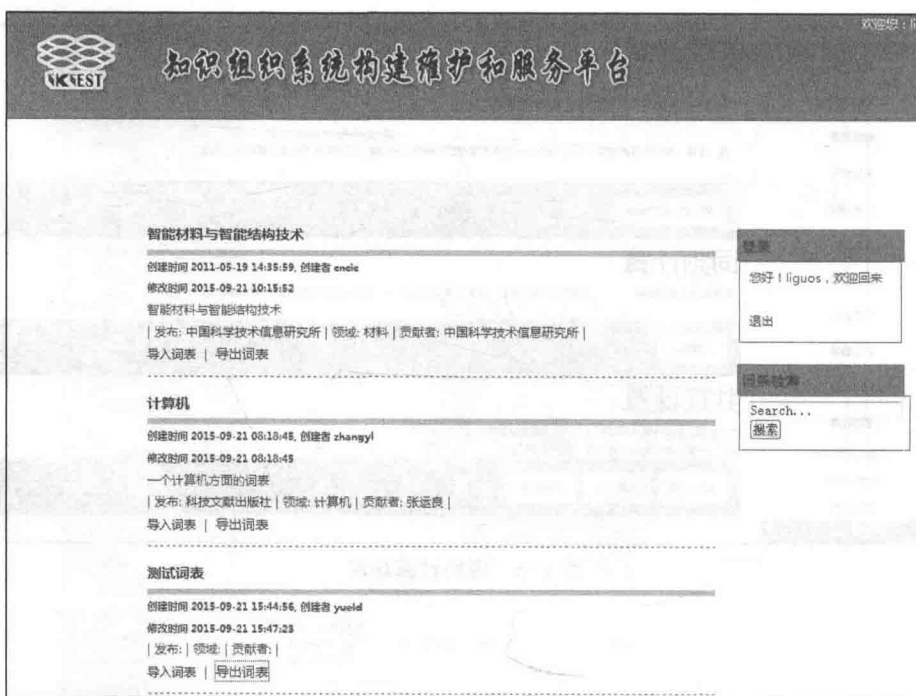


图 1-5 平台同时管理多个领域词表页面

平台支持多种分类法的管理,每一个领域词表可以同时支持多种分类法,一个分类法也可以运用到多个领域的词表中;运用分类法管理,增加分类法链接,可以为系统或者某部词表增加、修改和删除分类法的功能;在分类法列表,点击某部分类法的名称,可以进入该分类法的分类类目管理页面,在此页面可以对分类类目进行增加、删除和修改操作(如图 1-6 和图 1-7 所示)。

图 1-6 分类法管理

图 1-7 分类表实例

汉语科技词系统协同构建平台能够提供友好的用户界面,不但使普通浏览用户能尽快定位到自己感兴趣的信息,而且使后台编辑用户的工作量最大限度地减少;系统具有良好的运行效率,能够达到提高生产率的目的;系统应有良好的可扩展性,可以较容易地加入其他系统的应用;平台的设计具有一定的超前性、灵活性,能够适应企业生产配置的变化。本系统重点实现对词条的加工及信息展示,支持多批次数据处理,包含数据导入、词条管理、关系管理、属性管理、分类管理、定义管理、数据导出和词料计算等功能。

汉语科技词系统协同构建平台为各级用户均提供基本的可视化词系统构建界面,方便用户浏览、导入和导出词表,协助用户完成选词、词间关系推荐和构建等基本操作。系统的操作简单,清晰,提供交互操作的功能,各个功能系统展示清晰,便于用户操作。例如,检索结果的分页浏览及对检索结果的排序处理。

汉语科技词系统协同构建平台在词表管理中增加了词表的导入、导出功能,系统支持 CSV 格

式的文件导入、导出,导出的类型包括包含词条、包含词条关系、包含词条属性、包含词条定义及包含所有。在系统的统计信息中增加了统计汇总功能,根据词表统计词表中的词条、关系、属性的数量,同时统计3种类型的草稿、候选、审核状态,以饼图形式显示各类型所占的比例。

此外,汉语科技词系统协同构建平台在用户管理方面进行了完善。目前的用户主要包括两类:一类是普通浏览用户,既支持相关领域的专业人士,也支持一般的使用者;另一类是管理用户,这一类用户基本上都是对本领域的知识比较熟悉的专业人士,他们的主要职责是增加或维护系统的数据内容。用户管理是系统重要的一个部分,设置合理的用户组权限,有利于系统的功能分配。平台对管理员具有的管理权限、知识加工用户权限和普通用户权限重新进行了设计,使得权限的分配更合理。在用户管理功能里,管理员可以对系统中的所有用户的权限进行分配。用户管理主要有以下几个方面的管理:添加新用户、编辑用户、删除用户、增加用户权限、修改与删除用户权限(如图1-8所示)。

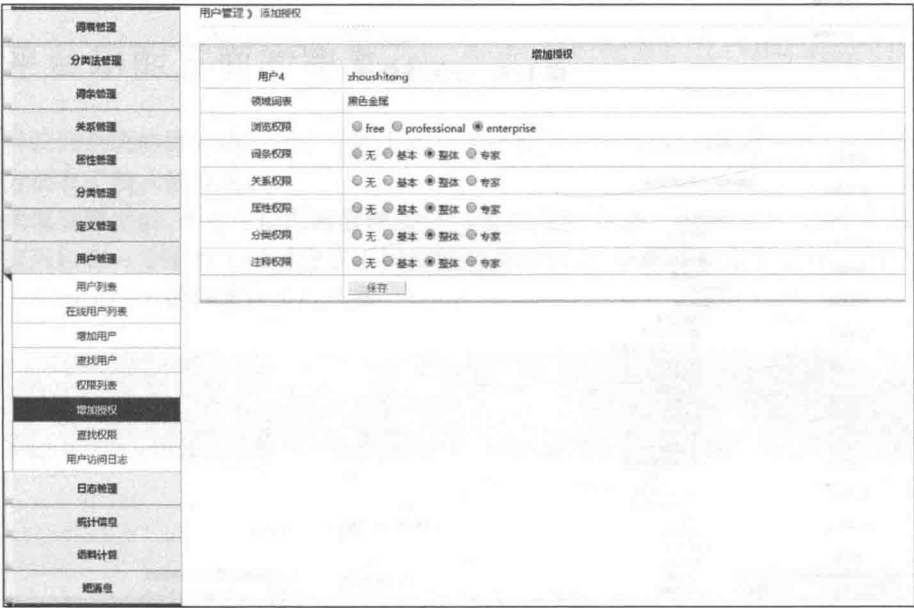


图 1-8 用户管理及用户权限修改功能

- 汉语科技词系统协同构建平台目前具有的功能分为前台和后台两部分。
- 其中,前台具备的主要功能如下:
- ①用户注册:实现了输入用户名、输入手机号和输入邮箱的功能,以及输入密码、确认密码和选填的功能;还实现了验证码、选择用户类型、注册信息和取消注册的功能;
 - ②登录:实现了输入用户名、密码登录系统的功能,以及取消登录和新用户注册功能;
 - ③主页:实现了输入查询信息进行查询的功能,以及查看词表信息的功能;
 - ④词条浏览:实现了根据分类法、词条首字母浏览词条的功能,以及查看词条详细信息、关系、属性等信息的功能。
- 后台具备的主要功能如下:
- ①词表管理:实现了词表的增加、浏览、删除、修改功能,以及设置默认词表功能;还实现了导入、导出词表功能。

②分类法管理:实现了分类法的浏览、增加、修改、删除功能,以及所属词表的设置功能。

③词条管理:实现了词条的浏览、增加、修改、删除基本信息功能,以及词条的草稿、候选、审核状态的修改功能和删除词条的管理;词条增加关系、属性、定义、分类描述的功能,还实现了按分类浏览词条及词条加工概况浏览的功能。

④关系管理:实现了词条关系的修改、删除、浏览功能,以及词条关系的草稿、候选、审核状态的修改管理;还实现了删除的词条关系的管理,以及关系类型的增加、修改、删除功能和逆关系类型的增加、删除等功能。

⑤属性管理:实现了词条属性草稿、候选、审核状态的修改功能和删除状态的词条属性管理功能;还实现了属性类型的增加、修改、删除功能,还实现了查找词条属性功能。

⑥分类管理:实现了词条分类草稿、候选、审核状态的修改功能,以及删除的词条分类管理功能和查找词条分类的功能。

⑦定义管理:实现了词条定义草稿、候选、审核状态的修改功能,以及词条删除状态的管理和查找词条定义功能。

⑧用户管理:实现了系统的所有用户的增加、修改、删除的功能,以及查看在线用户,查找用户功能;还实现了用户权限的增加、修改、删除功能,以及用户访问日志的查看、删除功能。

⑨日志管理:实现了系统中的词表日志、分类法日志、词条日志、词属性日志、词关系日志、词注释日志、词分类日志、用户日志的查看和删除功能。

⑩统计信息:实现了系统数据的统计浏览功能、工作质量详细统计功能、工作量详细统计功能,以及统计汇总功能;还实现了统计汇总功能中图形化展示的功能。

⑪语料计算:实现了语料的清空与导入功能、数据计算功能,以及关键词词频统计、统计结果的导出功能;还实现了关键词共现计算、结果导出功能。

⑫短消息:实现了短消息的查看、删除功能,发件箱消息的查看、删除功能,以及写消息功能和联系人查看、增加、删除功能。

3 汉语科技词系统应用探索

3.1 新材料领域

3.1.1 背景

制造业是国民经济的主体,是科技创新的主战场,是立国之本、兴国之器、强国之基,而材料是所有制造业的基石。新材料已成为高新技术的重要组成部分,渗透到国民经济、国防建设和社会生活的各个领域。2015年,国务院印发了《中国制造2025》,明确提出将新材料作为我国实现制造强国的重点发展领域。

科技情报对科技发展与创新意义重大,其中科技文献调研是获取科技情报的一种重要方法,对新材料领域进行科技情报调研是我国发展新材料领域的基础,从数量巨大且日益快速增长的新材料领域文献中获取准确、及时、快速、全面和简短的科技情报任务是一项迫切的任务。基于知识组织系统的新材料领域应用示范旨在以新材料领域的知识组织系统为基础,从大量的文献中快速、准确地分析出行业的发展趋势、各种技术的成熟度状况、各个主题的演化趋势等,为调研人员进行文献调研提供必要的技术支撑。

3.1.2 知识组织系统的基本情况

项目组采用了多种自动化方法组合从相关词系统、论文、报告、百科等抽取新材料方面的词汇、关系、定义等信息,然后对抽取的知识进行归类、组合,在此基础上投入人力对相关知识进行加工,完成了新材料词系统的建设工作。

目前,建设的新材料词系统涵盖新能源材料、纳米材料、石墨烯、压电材料、超导材料、高分子材料等领域,共包含词条11 738条,其中核心词条不少于2286条,基础词9452条,知识总量10 393条(包括定义472条、关系2749条、属性3066条、中英对译1820条、分类信息2286条)。

3.1.3 应用实践及效果

知识组织系统在新材料领域的应用示范中的作用主要体现在:①依据知识组织系统的科技文献的重新组织:传统的以中图分类法为基础的文献组织是一种粗粒度的组织方式,在新材料领域的应用示范中我们尝试了以词系统为基础的这种更细粒度的文献组织方式;②以知识组织系统为基础的科技文献数据库的检索及导航:在文献数据库检索时,不仅采用了传统的输入提示的方式,还可以以词系统为基础,进行用户输入检索词的更宽泛(上位词、下位词等)的输入提示;③以知识组织系统为基础的科技文献的主题分析、主题演化趋势分析及技术成熟度分析:在构造相应的分析模型时,我们对相应的概率主题模型进行了改进,将词系统加入进概率主题模型的构造过程中,通过实验证实了该方法的有效性。

目前,新材料领域应用示范系统的主要功能有:①学术论文的搜索;②科研人员的搜索及学术名片的展示;③学术期刊的搜索及学术名片的展示;④科研机构的搜索及学术名片的展示;⑤研究领域的技术趋势分析及主题演化趋势分析。系统已经部署到钢铁研究总院相应的平台上并有了一