

非物质文化遗产本体构建与语义化组织研究*

董坤^{1,2}

(1. 武汉大学信息资源管理学院, 武汉 430072; 2. 中南民族大学图书馆, 武汉 430074)

摘要: 针对当前非物质文化遗产分类组织方法的不足, 提出一个基于关联数据的非物质文化遗产语义化组织框架。通过构建非物质文化遗产本体描述模型, 实现非物质文化遗产知识元及其关联关系的语义化描述, 在其基础上基于关联数据发布技术实现了非物质文化遗产知识内容及内容之间关联关系的语义化整合与组织。

关键词: 关联数据; 本体; 非物质文化遗产

中图分类号: G201

DOI: 10.3772/j.issn.1673—2286.2014.10.008

联合国教科文组织在《保护非物质文化遗产公约》中将“非物质文化遗产”的概念界定为“被各群体、团体、有时为个人视为其文化遗产的各种社会实践”。我国作为世界上拥有非物质文化遗产数量最多的国家, 对濒临消失的非物质文化遗产进行抢救性保护一直是我国非物质文化遗产保护工作的重要内容。然而从信息组织层面看, 我国非物质文化遗产相关资源的数字化组织整理工作还存在着一定的不足, 主要表现在文献组织方法较为传统, 主要依据传统的分类组织方法, 如主题分类、等级分类、地域分类等组织方法, 这种基于非遗某一特征进行的单线索的线性组织, 无法反映非遗之间的固有联系, 无法表达非遗表现形式之间以及其子元素之间存在的复杂的关系, 造成查询与利用的不便。为有效解决这一问题, 本文针对我国非物质文化遗产的特点, 在借鉴语义网知识组织方法的基础之上, 设计了一个基于本体构建的非物质文化遗产知识组织模型, 基于该组织模型实现对我国非物质文化遗产组成因素、各因素之间关联关系的统一的语义描述与揭示, 为非物质文化遗产的语义化组织提供一种有效的方法, 最后基于关联数据发布技术进行了非遗资源的语义化组织系统实现。

1 非遗资源语义化组织现状

近年来, 为有效保护与宣传非遗资源, 实现保护工作的数字化与现代化, 在地方文化管理部门的推动下, 一大批非物质文化遗产数据库纷纷建立, 从现有非物质文化遗产数字化保护成果来看, 对非遗项目的组织模式普遍采用分类方法^[1,2], 对非物质文化遗产按其遗产地域、所属分类、保护级别等特征进行单线索的组织划分, 这种组织方法虽然具有较好的局部条理性, 但对非物质文化遗产外延下所蕴含的大量丰富的关联知识, 如非遗项目所包含知识内容本身以及非遗项目与项目之间、非遗项目与所处时空之间、非遗项目与传承人之间、非遗传承人之间、非遗项目与使用器具之间丰富的语义关联关系缺乏有效的揭示手段, 一方面造成非遗项目所蕴含的知识内容的极大流失, 另一方面由于缺乏统一科学的非遗组织体系, 已有非物质文化遗产数据库平台内容分散为一个个独立的信息孤岛, 难以在现有组织架构基础上实现对非物质文化遗产数字化保护成果的整合。

本体作为“共享概念模型的明确的形式化规范说明”, 通过领域本体的构建将该领域中的各种概念以及概念之间的关系进行显性、形式化的表现^[3], 为解决

*本研究得到湖北省高校图工委2012年科研基金项目“基于关联数据的高校知识管理体系构建研究”(编号: 2012ZD03)、中南民族大学中央高校基本科研业务费专项资金项目“基于关联数据的高校知识地图构建研究”(编号: CSQ14025)和2014年国家社科基金青年项目“高校图书馆嵌入式学科服务模式与质量评价研究”(编号: 14CTQ006)资助。

目前非物质文化遗产传统组织方法的缺陷,将非物质文化遗产领域所包含全部知识以及其语义关联进行系统、全面的揭示奠定了方法论基础。同时,本体支持逻辑推理的特性也为上层的基于语义的各种知识挖掘与发现提供了底层数据支撑。因此,将本体技术引入到非物质文化遗产组织领域无疑为非遗资源科学有序的组织指引了方向。目前,国内外还没有一个有影响力的应用于非物质文化遗产领域的特定本体,但近年来已有部分学者开始探索基于国内实际适合中国国情的非物质文化遗产本体构建,程齐凯、周耀林^[4]等将非遗资源划分为静态概念、动态概念与空间概念三个部分,并以此为基础构建了一个以非遗实体、静态实体、动态实体、空间实体与行动个体为顶层概念并以此不断继承细化的由28个概念实体所构成的四级非物质文化遗产本体框架,然而该框架对实体所指代的非遗资源外延及内涵的界定以及实体之间的关联关系没有进行深入探讨。郝挺雷^[5]定义了一个传统节日知识本体,并以国家非物质文化遗产端午节为例进行实证研究,该本体由事件、时间跨度、地点、行动者、物理实体和概念对象等六个核心实体以及属性构成,但该本体设计仅针对传统节日这一特定类型。从现有研究成果看,在对非遗资源知识元及其语义关联进行系统、全面、语义化的揭示与组织研

究仍然处于起步阶段。

2 非物质文化遗产语义本体构建

非遗作为文化形式和文化空间的有机结合体,非遗的管理不仅要关注其表现形式,更要关注其所生存发展的时间、空间、生态环境等因素。为实现对非遗知识元及其关联关系的语义化描述,首先需要构建一个元数据本体实现对非遗领域内的核心概念以及概念之间的语义关系进行精确的语义化描述。目前,在非遗的领域内还没有出现特定本体,但在其上层的文化遗产领域,已经出现了一些具有一定影响的本体方案,如CIDOC CRM^[6]、ABB、AAT等,其中CIDOC本体基于面向对象的模型,描述关于文化遗产的概念以及概念之间的关联关系,其总共包含62个实体类,148个属性定义,对文化遗产具有最大的描述能力和灵活性。本研究在CIDOC CRM本体的基础,结合vcard、Geoname等本体词汇库并进行适当扩展,譬如构建ICH-Project类定义非遗项目,扩展定义ich:inheritor属性定义非遗项目与传承人之间关联关系,ich:inherit from属性定义非遗继承人之间继承关系,ich:accepted date属性定义非遗项目申报时间,本研究所构建的非遗元数据本体如图1所示。

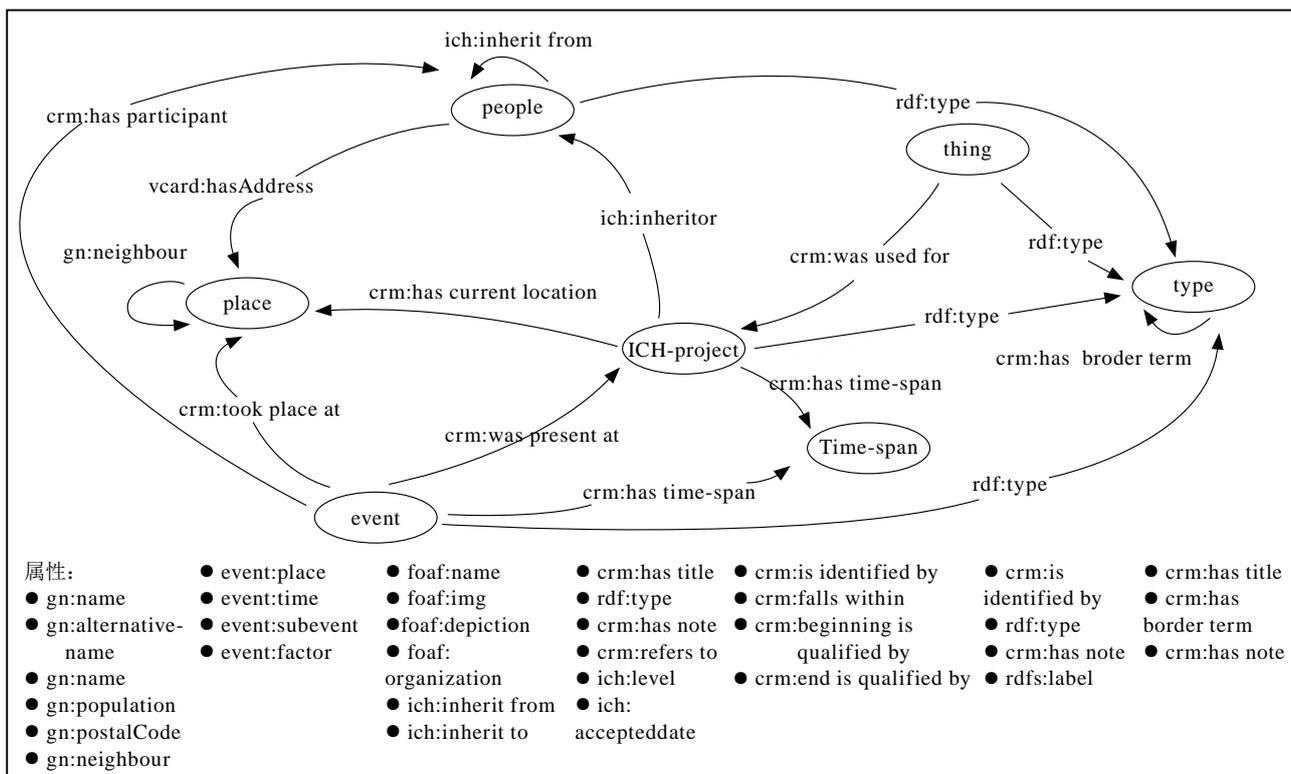


图1 非遗元数据本体

非遗元数据本体由非遗项目 (ICH-project)、人物 (Person)、位置 (Place)、事件 (Event)、时间段 (Time-span)、类型 (Type)、事物 (Thing) 等六个核心类组成,实例之间关联关系通过类属性来揭示。

非遗项目 (ICH-project) 类实体为我国非物质文化遗产的知识表征,其描述属性项主要包含我国非物质文化遗产的相关特征,如名称、别名、内容描述、类型、入选名录等信息。

人 (Person) 是非物质文化遗产所处人文环境的重要组成部分,通过人的作用,非物质文化遗产才能够得以产生、传承并得到研究、推广。在某种程度上,人是我国非物质文化遗产的唯一的依附载体。在本模型中,人物类是与非物质文化遗产相关联的一切人物的抽象,可以是非物质文化遗产的创造者、传承者,也可以是其推广者、研究者,其基本属性包括人物的姓名、出生年月、个人简历、相关图片等信息。

类型 (Type) 实体主要应用于非遗知识元素分类体系的构建,其类别实例可以依据不同的分类准则进行定义,可以来源于各类受控词表,也可以依据不同的分类体系来定义,如依据我国《非物质文化遗产名录》将非遗分为民间文学、民间音乐、传统医药、民俗等十大类,依据非遗的民族特性又可划分汉族、土家族、苗族等类别。类别实例之间还可以利用类属性 `crm:has border term` 进行层级关系的搭建,建立一个多视角、多层次、灵活的分类体系,其他类实体可分别构建与类型实体的映射关系。

时间段 (Time-span) 主要表征特定的时间范围,类属性包括时间段的起始时间、结束时间等。时间段类主要用来描述非遗项目在时间维度演化发展的特定窗口,如非遗的起源时期、发展时期、非遗项目申报时间等。

事件 (Event) 类指代一切与非遗项目有关的活动,如祭祀活动、集体仪式、文艺活动、竞技游戏等。非遗项目的生存、发展与其所处特定环境的特定民俗活动有着极为紧密的联系,如被誉为中国戏剧活化石的“恩施傩戏”、民间舞蹈“土家族撒叶儿嗬”与民间祭祀活动之间关联的关系、民俗“端午节”与“赛龙舟”活动的关联关系等^[7]。

非物质文化遗产存在着较为鲜明的地域特征,同一地域相同的语言文字、风俗习惯、文化、历史源流往往孕育着反映该地域不同社会侧面的非遗内容,同一地域或相邻地域的非遗项目之间往往存在有着不同程

度的关联关系。在 ICH 模型中位置 (Place) 实体一方面对非物质文化遗产所流传的区域进行描述,同时也为非物质文化遗产的发现提供多样性的捕获和发现手段。在本文中基于 Geoname 本体库对非物质文化遗产所处地域基于其地理关系进行语义化组织。

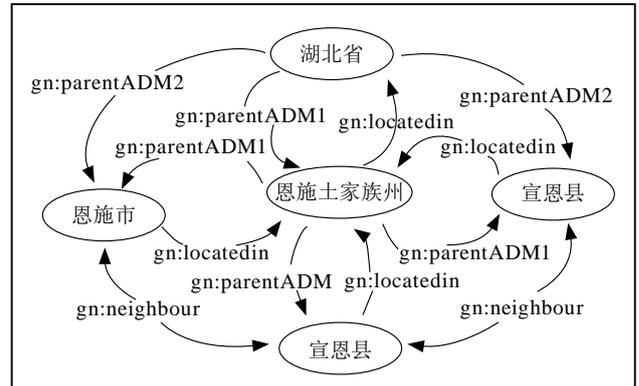


图2 位置实体语义关系示意图

非物质文化遗产也有其物质化的一面,大量的非遗项目需要依靠物质层面的事物来表达和呈现。如傩戏面具作为非遗项目傩戏造型艺术的重要手段,是傩戏区别于其他戏剧的重要特征,此外还有大量展示非遗项目内容的影、音、文字等文献资源。事物 (Thing) 类实体主要揭示与非遗相关的物质元素。

类之间的关联关系通过类属性来定义。部分核心类属性如表1所示。

3 基于本体的非遗资源语义化组织功能实现

3.1 非物质文化遗产语义组织框架构建

在非遗元数据本体描述模型的基础上,本研究提出非遗资源语义整合框架,基于非遗本体从静态、动态、时空等视角将非遗表现形式与其所处的文化空间知识内容进行整合,并利用关联数据发布技术实现对非物质文化遗产丰富知识元及其关联关系的发布与展示。该框架由四层结构组成,如图3所示,从下至上分别为源数据层、语义层、数据关联层与应用层。

源数据层是非物质文化遗产所涉及知识内涵的最初始状态的信息表达形式,如文档、互联网页所记载的非遗的表现形式、所处地域、相关人物、发展时间、相关文献等信息。源数据层为上层的非遗知识提取、知识

标引、知识组织功能的实现提供基础数据支持。

语义层基于语义本体元数据模型,实现本体资源的RDF化。实体RDF化是非遗资源语义化组织的关键一环,实体RDF化需要相应本体的支持^[8]。语义层基于本文所提出的非遗本体概念模型对底层源数据层进行知识与关系的抽取并转换为RDF格式的语义元数据。

数据关联层基于关联数据发布的技术,将基于语义本体所构建的RDF元数据在语义层上相互关联起来,将非遗相关资源发布为一个有机联系的统一的整体,每个资源都可通过HTTP协议进行直接访问,并可沿着RDF链接访问其他资源,向上层提供一个统一的富含语义的知识视图^[9]。

应用层基于数据关联层的知识组织架构,开发实现上层的各项语义应用服务功能,如语义检索、可视化语义知识地图、知识推理与发现等应用服务。

3.2 非遗资源语义化组织系统实现

本研究将以湖北、湖南两省非遗项目为例,基于上文所提出的非遗资源语义组织框架及非遗元数据本体,将湖北、湖南两省非遗项目所蕴含的知识内容以关联数据的形式进行发布,在其基础上实现非遗知识基

表1 部分属性

Properties	Domin	Range	说明
ich:inherit from	People	People	非遗项目传承人之间传承关系
ich:inheritor	Ich-project	Person	非遗项目与传承人之间的对应关系
crm:has current location	Ich-project	Place	非遗项目与流传地域之间的位置关系
vcard:has Address	Person	Place	人与地域之间的位置关系
rdf:type	Ich-project	Type	非遗项目与类别之间的分类关系
crm:has broder term	Type	Type	不同类别的层级关系
crm:has time-span	Ich-project	Time period	非遗项目与时间之间对应关系
crm:was present at	Ich-project	Event	非遗项目与事件之间的关联关系
crm:motivated	Ich-project	Event	非遗项目与事件之间的驱动关系
crm:has participant	Event	Person	事件与参与者之间的关系
crm:took place at	Event	Place	事件与事件发生地域之间关系
Crm:was used for	Thing	Ich-project	事物与非遗项目之间的被使用关系

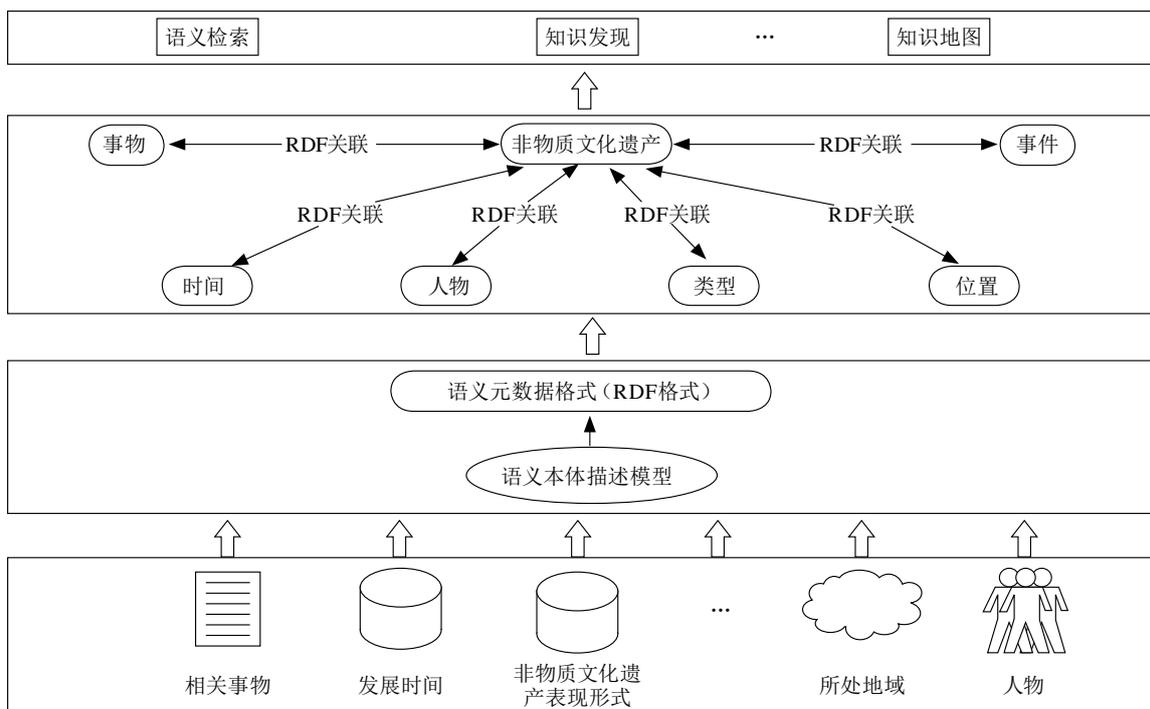


图3 非物质文化遗产语义组织框架

于语义的浏览、检索功能。

本研究共整理出湖北、湖南两省国家级、省级、市级非遗项目内容321项,包含非遗内容介绍、地域信息、传承人信息、相关事物信息等,采用在线映射方式实现非遗数据的RDF化及关联数据的发布。因为非遗信息数据量较大,内容需要不断更新,因此所采用的具体方法是将非遗信息借助于关系数据库进行存储,基于D2RQ平台^[10]实现关系数据库与RDF关联数据的语义映射,将关系数据库发布为关联数据并提供语义查询检索接口。

D2RQ基于一个可个性化定制的D2RQ mapping file将关系数据库中数据直接映射为语义网RDF格式并建立对应的RDF链接关系。D2RQ提供RDF数据的浏览和检索接口,以供上层的RDF浏览器、SPARQL^[11]查询客户端或传统的HTML浏览器调用。

发布为关联数据的非遗知识浏览界面如图4所示。



图4 非遗信息的关联数据浏览

用户可分别点击页面上部Event、ICH-Project等类名称标签浏览类实例数据目录,点击目录链接即可查看实例数据内容。同时,用户也可以以任一实例数据为起点沿着关联数据所构建的RDF语义数据链,从任意一个非遗知识节点开始不断发散式访问浏览其他相关非遗知识内容,如从非遗项目目录→非遗项目→非遗传承人→传承人相关文献资源,非遗项目→流传地域→临近地域→非遗项目→相关事物等。

基于D2RQ平台所提供的SPARQL关联数据语义查询接口,用户可直接编辑SPARQL查询语言,根据系统所构建的语义关联检索所需信息。非遗项目“龙凤书”的类实例SPARQL查询检索页面如图5所示。



图5 基于SPARQL的语义检索接口

4 结语

本文提出了一种基于关联数据的非遗知识语义化组织方法,基于所构建的非遗语义本体,实现了对非遗资源及其关联关系的语义化揭示与组织,最后完成了一个原型系统的构建。该方法相对于传统分类、主题等组织方法的优点是:

(1) 改变了传统分类组织、主题组织等单线索式的组织模式,实现了对非遗知识基于其表现形式、人物、地域、相关事件、相关事物的多线索的有序管理,弥补了传统的单线索的分类组织方法功能上的缺陷。

(2) 基于RDF Dump协议或SPARQL协议外界能够便捷获取系统内的知识内容,同时也能将外部的非遗知识数据灵活地与系统内数据进行关联整合,可形成一个可无限扩展的非遗知识网络,向用户提供一个统一的蕴含丰富非遗知识元及其语义关联关系的知识视图。

(3) 基于RDF和非遗语义本体强大的语义描述能力,对非遗领域丰富的知识元以及知识元之间的语义关系进行了充分的揭示,为上层非遗领域的专家系统、知识推理等语义服务奠定了坚实的基础。

在后续研究中,将在非遗资源的语义化组织基础之上进一步研究基于关联数据知识组织系统的非遗资源的知识挖掘,例如基于规则库等方法的非遗知识推理与挖掘方法,同时构建基于自然语言的语义知识检索系统,让系统更加实用化。

参考文献

- [1] 湖北省非物质文化遗产资源库[EB/OL]. [2014-10-19]. <http://hbqxc.library.hb.cn/Article/hbfwz/Index.htm>.
- [2] 陕西省非物质文化遗产资源库[EB/OL]. [2014-10-19]. <http://www.snwh.gov.cn/feiwuzhi/>.

- [3] 杨建林. 基于本体的文本信息检索研究[J]. 情报理论与实践, 2006(5):598-601.
- [4] 程齐凯, 周耀林, 戴旻. 论基于本体的非物质文化遗产分类组织方法[J]. 信息资源管理学报, 2011(3):78-83.
- [5] 郝挺雷. 传统节日知识本体的构建及表示方法研究[D]. 武汉: 华中科技大学, 2011.
- [6] The CIDOC CRM [EB/OL]. [2014-07-15]. <http://www.didoc-crm.org>.
- [7] 谭志国. 土家族非物质文化遗产保护[D]. 武汉: 中南民族大学, 2011.
- [8] 沈志宏, 刘筱敏. 关联数据发布流程与关键问题研究[J]. 中国图书馆学报, 2013(3):53-62.
- [9] Link data-Connect distributed data across the Web [EB/OL]. [2014-07-15]. <http://linkeddata.org/>.
- [10] D2RQ-Accessing Relational Databases as Virtual RDF Graphs [EB/OL]. [2014-07-10]. <http://d2rq.org>.
- [11] SPARQL 1.1 Query Language [EB/OL]. [2014-07-10]. <http://www.w3.org/tr/sparql11-query>.

作者简介

董坤, 男, 1979年生, 武汉大学信息资源管理学院博士生, 中南民族大学图书馆馆员, 研究方向: 数字图书馆, E-mail: spevan@126.com。

Research on Intangible Cultural Heritage Organization

DONG Kun^{1,2}

(1. School of Information Management, Wuhan University, Wuhan 430072, China; 2. Library of South-Central University for Nationalities, Wuhan 430074, China)

Abstract: Aiming at the disadvantages of the current classification organization methods of intangible cultural heritage, the paper brings forward a semantic organization construction of intangible cultural heritage based on linked data. Through the construction of ontology model of intangible cultural heritage, we realize the semantic description of knowledge elements and associations between them. On the basis of the ontology, the integration and organization of intangible cultural heritage based on linked data come true.

Keywords: Linked data; Ontology; Intangible cultural heritage

(收稿日期: 2014-09-30)

■ 书 讯 ■

《中国高被引分析报告2012》

在您所关注的研究领域, 哪些论文最受同行关注? 哪些研究主题最为热门? 哪些学者最具学术影响力? 哪些期刊最获同行认可? 又有哪些机构占据领域研究的制高点? 上述问题请您参考近期出版的《中国高被引分析报告2012》。

该书将理、工、农、医、人文、社科等领域划分为51个学科, 综合分析了各个学科的研究热点与前沿、高影响力论文、高影响力作者、高影响力期刊和高影响力科研机构, 并以关联图谱的方式展现各种学术关系, 有助于科研人员及时发现并跟踪研究热点, 可为科研管理机构评估科研能力提供依据, 还有利于期刊编辑部监测本刊学术影响力, 是高等院校、科研院所及期刊编辑部等相关单位和人员参考工具书。

该书以“中国知识链接数据库”为依托, 数据样本覆盖我国6000余种期刊的论文及引文, 分学科揭示高影响力的学者、研究机构(大学、研究所、医院等)、地区(省/自治区/直辖市)、学术期刊、图书、外文期刊和会议录, 并采用共词分析、共被引分析和合著分析等方法绘制出各学科的前沿主题分布以及作者、机构和期刊间关联的知识图谱。

《中国高被引分析报告2012》由中国科学技术信息研究所编著, 科学技术文献出版社出版, 全书86万字, 定价298.00元。