

·智库研究·

# 基于语义关联的智库资源知识组织研究\*

李慧佳<sup>1,2</sup> 王楠<sup>1</sup>

(1.中国科学院西北生态环境资源研究院兰州文献情报中心 甘肃兰州 730000)

(2.中国科学院大学 北京 100049)

**摘要:**智库是知识决策咨询体系的重要组成部分,其主要通过智库资源的知识发现和智库专家的经验智慧相结合来生产智库产品,进而影响相关领域的知识决策。利用语义关联来构建的智库资源知识组织体系可实现智库资源的知识化、有序化、关联化组织,从而实现智库资源的利用、共享、传播和增值等知识化服务。

**关键词:**智库;语义关联;知识组织;CSpace DR;知识服务;知识决策

中图分类号:G255 文献标识码:A DOI:10.11968/tsyqb.1003-6938.2020015

## Study on Knowledge Organization of Think Tank Resources based on Semantic Link

**Abstract** Think tank is an important part of the scientific and technological strategic decision-making consulting system, through the vast amount of knowledge resources and experience and wisdom to influence the relevant areas of knowledge decision-making. This paper constructs knowledge organization system framework of think tank resources with semantic link, so as to realize the knowledgeable, ordering and linkable of think tank resources, and then realize the use of think tank resources, sharing, dissemination and value-added knowledge services.

**Key words** think tank; semantic link; knowledge organization; CSpace DR; knowledge service; knowledge decision

在迈向知识决策的道路上,智库占据重要的地位。智库通过海量的知识资源和智囊团的经验及智慧来影响相关领域的知识决策,因此由专家、机构、资讯、智库出版物、专题数据库等组成了智库这一共同体。在以知识创新引领创新驱动发展的战略背景下,智库作为决策咨询体系的重要组成部分,针对不同领域构建基于语义关联的智库资源知识组织体系,一方面可为创新型的科技战略研究、科研绩效分析、科技与产业竞争力分析等科技决策与咨询工作提供流程框架,最终形成以各领域语义关联知识资源发现组织、集成和深度分析能力为基础,有效开展面向领域科技决策需求的战略情报研究与智库服务能力体系,另一方面可为决策者、智库专家等用户提供精准的智库资源服务。

### 1 资源知识组织是智库决策的有效支撑

近几年,大数据时代的到来和智库数量呈现飞跃

式增长的趋势,使得决策者和智囊团专家陷入信息泛滥、信息无序的困境中,大量过剩的信息反而成为智库决策束缚。迫切需要能够将智库资源进行科学处理、组织、关联,最终形成真正可以解决现实问题的知识资源。

#### 1.1 从元数据到语义关联:不断发展的知识组织方法

在知识计算等技术的推动下,知识组织的方法、工具从元数据、本体模型、知识地图,到语义关联,得到了不断的继承和发展<sup>[1]</sup>。其中,基于元数据的知识组织方式主要擅长揭示文献资源整体结构,但无法实现异构资源之间错综复杂关系的揭示;基于本体模型的知识组织主要是通过词表、语义关联技术、领域本体等方法来构建知识组织框架,是目前较为流行的知识组织方法之一。如JeromeDL项目通过构建MareOnt本体来实现各种元数据的语义转换<sup>[2]</sup>。但此种方法受到资源质量参差不齐、学科领域差异性,以及缺乏可

\* 本文系国家自然科学基金青年项目“基于语义关联的专题学术资源遴选方法研究”(项目编号:19CTQ007)、甘肃省科青年科技基金计划项目“基于名称规范的甘肃省科技资源服务平台建设”(项目编号:17JR5RA321)、中科院兰州文献情报中心创新能力项目“全球智库信息集成服务系统”(项目编号:Y8AJ012016)与中国科学院“西部之光”青年学者A类项目“基于学术大数据的专题化信息自动采集与组织技术研究”(项目编号:Y9AX011001)研究成果之一。

收稿日期:2019-12-15;责任编辑:胡刚

用的互操作性标准和规范等方面的因素的制约,导致本体模型在构建知识组织的可操作性较差。以知识地图的方法进行学术资源的知识组织,主要是通过可视化的形式来展示资源内部和外部间的关系,优点是可以更加直观、便捷的呈现知识资源,从而实现了知识的重用、共享和创新<sup>[3]</sup>。但目前国内外关于知识地图以理论研究为主,而知识地图的构建受到应用领域的局限,缺乏通用性和普适性,很难展开全方位的应用及研究。

语义关联通过识别和分析数字资源的语义信息,并通过自动整合、转换与处理,来解释不同资源之间的语义关系<sup>[4]</sup>。利用语义关联进行数字资源的知识组织可以弥补元数据方法只能用于表示资源语法关系的缺陷,并借助本体模型来表示资源的语义信息,从而改变数字资源知识组织的树状结构,通过资源见得深度关联形成无限延伸的网状结构,打破了本体模型在应用领域内的局限,为不同学科领域、多源异构数字资源的知识组织提供了技术支持。

## 1.2 国内外资深智库的资源知识组织主要策略

智库的作用是决策支持,但智库的资源范围、数量、质量、服务内容、方式等都直接影响决策<sup>[5]</sup>。智库决策需要全面而准确的知识资源作为保障,国外资深智库都有可靠的信息来源,以及专门的情报信息网络。如斯坦福研究所在美国各地区设有分支机构,并且在欧洲、中东、日本和中国香港等地均设有办事处,这些构成了斯坦福研究所的情报信息网络<sup>[6]</sup>。但多种来源种类各异的情报信息网络必须通过有效的资源组织才能真正为智库服务。

目前,国内外智库的资源知识组织策略主要有:(1)按资源类型分类组织。智库机构将其直接生产或间接搜集的资源进行归类存储,这样的资源组织方式可以实现智库资源的简单快速整合和存储,但也存在很多弊端:分类标准不统一;元数据不规范;资源冗余较多;开放权益不明确;跨库检索不易实现等;(2)按专题数据库组织。随着信息技术的发展,越来越多的智库机构开始通过建设专题数据库的方式来组织智库资源。知名智库几乎都建有各具特色的智库资源数据库,如兰德公司开发了十几种专题数据库用于存储各类专题资源(健康调查数据库、统计

数据库、世界恐怖事件数据库等),这些数据库都是兰德公司智库决策的重要支撑<sup>[7]</sup>;瑞典斯德哥尔摩国际和平研究所(SIPRI)自建了5个专题数据库,分别是武器转让、国际关系与安全趋势、军火工业、多国和平行动、军事开支专题数据库<sup>[8]</sup>,但也存在着资源难以有效关联和发现等问题;(3)智库资源共建共享。根据美国宾夕法尼亚大学“智库与公民社会计划”(TTCSP)项目发布的《全球智库报告》可以看出全球智库从2008年的5465家增长到2018年的8162家,10年智库新增率为49%<sup>[9]</sup>。智库的不断增加使得相似的专题上会存在多个独立的数据库,从而导致信息数据重复存储,使用效率低下等问题。解决这一问题最好的办法就是实现智库资源的共建共享。智库资源的共建共享最早可追溯到1992年德国国际政治与安全研究所与德国12家研究机构共同建立的信息网络中心——EINRAS(European Information Network International Relations and Area Studies),该网络中心的任务是促进在研究和实践中收集、管理、分析和使用信息和文件的问题进行实际合作和交流<sup>[10]</sup>。EINRAS的建立可以最大程度的整合资源,确保相关领域的国际竞争力;同时,可以避免资源的重复建设,实现资源的最优配置。2002年,著名智库美国企业公共政策研究所与布鲁金斯学会合作共建了网络资源平台,解决了美国总统选举制度改革过程中出现的大量重要资料分散的问题<sup>[11]</sup>。

## 1.3 语义关联为智库的资源知识组织提供了新选择

语义关联不仅可以将错综复杂的智库资源有序呈现,而且可以揭示隐藏在智库资源深层、不易被人们察觉的隐形知识。因而,基于语义关联的智库资源知识组织体系构建的价值在于:

(1)智库资源知识化。智库资源知识化过程是通过对于智库资源的梳理、加工和再组织,从而形成可以实现咨询和决策的智库知识。智库资源知识化可以使智库突破传统智库服务模式,挖掘智库资源中的隐形知识,并将其转换为显性知识,从而提高智库资源的利用价值。

(2)智库资源有序化。智库资源有序化是对相互独立的智库资源按照一定的需求进行分类、关联和管理,更有利于智库知识的检索利用。相互独立的智

库资源本身无法体现知识,只有通过对智库资源的有序化处理,以形成更有价值的智库知识,从而改善和提升智库服务的质量。

(3)智库服务知识化。智库服务知识化是智库资源知识组织体系构建的最终目标,利用构建智库资源的语义关联模型,能从各种显性和隐形智库资源中有针对性的提炼出智库知识,最终实现智库知识的利用、共享、传播和增值等功能,从而为智库咨询、决策和战略规范提供有效支持。

## 2 智库资源知识组织的内容

### 2.1 智库资源的实体类型

智库资源主要包括智库机构、智库专家、智库产品、智库动态(见图1)。各类智库资源相互碰撞、有效融合,最终形成新的、增值的智库知识。

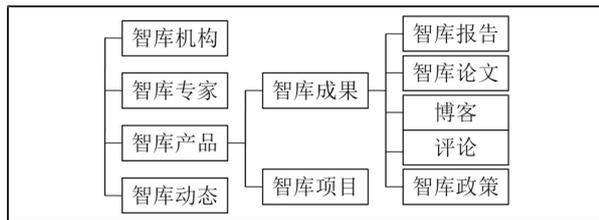


图1 智库资源实体类型

(1)智库机构。智库机构是为智库提供信息数据的主要载体,智库机构可以直接生产资源,也可以间接搜集信息和数据。因此智库机构是智库资源结构中处于顶端的资源,是其他智库资源的触发器。依据不同的智库构建理念可将智库分为学术型智库、契约型智库等,但无论何种类型智库都依托于一个或多个的机构成立。TTCSP发布的《全球智库报告2018》,通过对8000余家智库依托机构进行调研分析发现,智库依托机构以高校、科研院所、学协会、基金会、政府机构、企业为主。

(2)智库专家。智库专家是提高智库影响力的重要因素之一,也是智库的核心竞争力所在,是智库资源中不可或缺的宝贵资源。智库专家既可以是隶属于智库机构的专职专家,也可以是智库机构以外的领域专家。智库专家对信息数据具有超强的敏感性,在特定领域有一定的社会地位和话语权,具有洞悉前沿热点,把握政策趋势的能力,并能正确的对正在发生或未来会发生的事情做出判断,以及给出咨询

建议。国内外众多高影响力的智库都针对不同的领域设立不同的研究部门,用于组织相关智库专家集中为特定领域提供信息服务。如德国科学与政治基金会(SWP)针对其下8个专题研究部门设立了专门的信息研究室,用于组织专家为该基金会以及德国联邦议院和联邦政府各部门提供信息服务<sup>[12]</sup>。

(3)智库产品。智库产品是智库研究成果的展示,也是智库机构开展咨询和影响决策的主要方式之一<sup>[13]</sup>。智库产品的类型没有统一标准,大致可分为报告(Report)、论文(Article)、博客(Blog)、评论(Comment)。另外,还有智库机构(如学协会、基金会等)会面向特定地区发布智库项目,用以寻求特定研究主题的合作研究。一般而言可将智库产品分为两类:一是智库机构和智库专家直接产生的智库成果,包括智库报告、智库论文、博客、评论和智库政策。其中智库报告包含技术报告(Technical Report)、通用报告(Report)、年度报告(Annual Report)、工作报告(Working Paper)等;智库论文包括期刊论文、会议论文、灰色文献等;博客是指智库专家通过社交平台发表的非正式言论;评论是指智库机构或智库专家提出的具有评价性、针对性的观点或建议;智库政策不仅是正式出版或发布的政策性文档(Policy Paper),也有通过特定标题(Policy/Decision/Proposal/Recommendation)从智库报告中抽取得到的片段性文档。二是智库机构对外公布的智库项目,旨在通过多方合作来完成特定主体的研究。

(4)智库动态。智库动态是指在智库机构网站上发表简明的、公开的、具有时效性和针对性的文章或消息。智库动态可以使人们以最快的速度了解特定领域最新的或最具价值的信息。由于智库动态具有固定的结构,一般包括标题、导语、主体、背景和结语,可将其视为结构化文档。因此不仅智库动态本身是重要的智库资源,智库动态中还有大量隐形资源可为智库的政策制定提供支持。

### 2.2 智库资源的实体要素

以智库机构为主线出发,可将智库活动过程中的概念实体化,本文采用复用PROV模型中的Organization、Person、Activity、Asset实体来表达智库资源,即智库机构(prov:organization)、智库专家(prov:per-

son)、智库动态(prov:activity)和智库产品(prov:asset)。

(1)智库机构实体要素。作为智库资源的首要资源,智库机构是智库知识的源泉,主要包括机构名称、其他名称、机构简介、机构网址、成立日期、机构类型、语种、国家、电子邮箱、联系电话和联系地址等。

(2)智库专家实体要素。智库专家属包括专家姓名、个人简介、研究方向、出生日期、职称、国家、性别、头像和电子邮箱等。

(3)智库产品实体要素。智库产品的属性包括:DOI、题名、其他题名、摘要、语种、产品类型、格式、学科、出版日期、作者、出版者、URL和关键词等。在DC元数据标准或其他文献资源的元数据标准中没有适用于智库产品类型,可直接被复用的元数据字段,因此,对于智库产品类型的元数据将采用复用其他本体中实体要素的方式进行描述。

(4)智库动态实体要素。智库动态的属性包括:题名、其他题名、内容、语种、URL、作者、出版日期和关键词等。

### 3 基于语义关联的智库资源知识组织框架

基于语义关联的智库资源知识组织框架分为元数据层、关联层和应用层等三个层次(见图2)。该框架通过对智库资源进行元数据规范、语义化描述,进而提供统一的资源访问机制,实现多源异构资源之间的语义互操作。

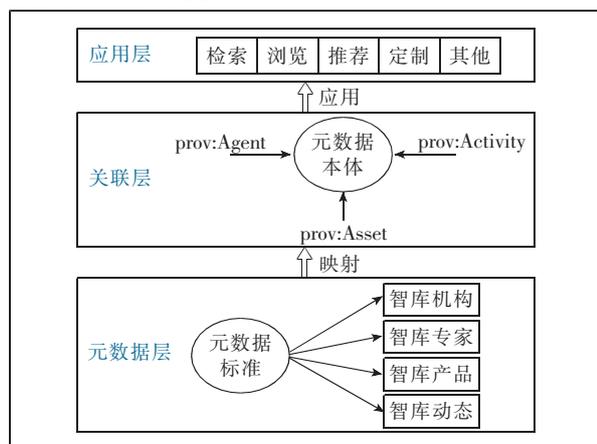


图2 智库资源知识组织框架图

#### 3.1 元数据层

资源组织/存储系统中最基础的组织方式就是通过元数据标准对资源进行规范化描述,常用的元

数据标准有支持网络资源的Dublin Core、IAFA Template、CDF、Web Collections;支持文献资源的MARC、Dublic Core。对于多源异构多种类型智库资源而言,往往会用多种元数据标准对资源进行描述。这些元数据标准之间既存在相似之处(如共享相同的核心元素),但并不完全兼容。为了使不同类型多源异构智库资源的元数据具备良好的兼容性和互操作性,就必须在元数据核心元素的基础上最大限度的整合元数据标准,形成一套适用于智库资源的元数据标准。然而,元数据虽然提供了一定的语义基础,使资源有了基本的微观结构,但是元数据并不能完全解决语义异构问题,包括资源采用不同元数据方案所造成的微观结构的异构问题以及资源对象之间存在的复杂的关联关系<sup>[14]</sup>。

#### 3.2 关联层

为了更有效呈现多源异构资源之间关联关系,需要在智库资源知识组织框架中元数据层的基础上构建智库资源的元数据本体以及关联关系,进而实现不同类型智库资源元数据的语义互操作。元数据本体的构建方法有两种:一是通过本体描述语言构建一个集成的元数据本体,并基于集成的元数据本体实现不同源数据的语义互操作;另一种是在通过本体描述语言对每一种元数据标准进行本体化描述,然后通过不同元数据本体之间的映射关系来实现不同元数据本体间的语义互操作<sup>[15]</sup>。智库资源是由网络资源和文献资源组成,并且智库资源的组成相对稳定,不会出现频繁增加新的元数据标准的现象。同时,DC元数据标准既可以用来描述网络资源,也可以用来描述文献资源。因此,在智库资源知识组织框架中构建一个集成元数据本体无疑是最优方案。首先,构建一个基于DC的核心元数据本体;其次,对于专门元数据元素可通过复用其他元数据标准中元素集的形式对核心元数据本体进行扩展;最终形成智库资源的集成元数据本体。

元数据本体解决了智库资源中文献类资源的元数据描述问题,同时揭示了文献资源之间的显性关系。为了实现不同类型、多源异构的智库资源之间的互联互通,需要在智库资源集成元数据本体的基础上采用关联数据方式对智库资源进行再组织。通过

关联数据方式对智库资源进行知识组织的优势在于用户既可更全面、完整的直接浏览和检索到相关知识,也可实现与其他相关资源、信息和数据的沟通。智库资源知识组织的关联层实现了智库资源的自动发现、采集、组织、存储等功能,并向应用层提供对外服务的开放服务接口。

### 3.3 应用层

智库资源知识组织框架的应用层是在集成智库资源的基础上,提供统一的服务接口,为用户提供增值服务,如智库资源语义检索、关联浏览、精准推荐、个性化定制等。语义检索是在传统词检索方式的基础上,进一步提供的问答检索方式;关联浏览是在语义检索的结果上为用户提供其他相似、相关资源的浏览;精准推荐是根据用户的检索行为将相似、相关资源主动推荐给用户的;个性化定制是用户可通过主题词、资源类型进行智库资源的定制,同时也可对智库研究热点、可视化分析的结果进行定制。

## 4 基于语义关联的智库知识组织实践

为了提高全球智库资源的组织的效率,中科院西北环境生态资源研究院兰州文献情报中心以 CSpace DR 为底层平台构建全球智库信息集成服务系统,结合元数据法和关联本体模型进行多源异构智库资源的知识组织体系构建。该服务系统旨在快速集成国内外权威智库资源,通过对智库资源(智库、智库专家、智库报告、智库观点、机构、新闻等)的知识组织采用本文设计的知识组织三层框架进行有序化组织和关联化呈现,从而实现了智库资源的利用、共享、传播和增值等知识化服务。本文主要对构建智库资源知识组织体系关联层的内容进行研究。

### 4.1 智库资源的实体-属性-关系本体构建

首先通过对不同类型多源异构的智库资源进行分析,并将分析得到的概念抽象为实体类,建立实体类之间的关系,从而构建组织关联智库资源实体类的模型,其中类和关系的表示参考复用了 PROV 本体、VIVO 本体的类和属性以及 DC 元素(见图 3)。

DC Term 中对实体关系的定义较为宽泛,在对多源异构智库资源的实体关系进行描述时不能明确表达实体关系的含义。因此,我们依然通过复用其他

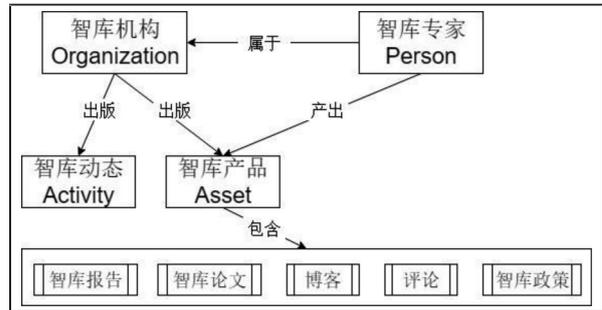


图3 智库资源实体关联关系图

本体模型中实体关系的方式对智库资源实体关系进行扩展描述,主要复用 prov 本体模型中的关系属性(如 pro: affiliatedOrganization、prov: generated)。

(1)属于(prov:affiliatedOrganization):复用 prov 模型中 affiliatedOrganization,用以表示智库专家(prov:person)隶属于智库机构(prov:organization);

(2)出版(dcterms:hasPublished):复用 dcterms 元素中的 Published,用以表示智库机构(prov:organization)出版或发布智库动态(prov:activity)或智库产品(prov:asset);

(3)产出(prov:generated):复用 prov 模型中的 generated,用以表示智库专家(prov:person)发表了相关的智库产品(prov:asset);

(4)包含(dcterms:hasPart):复用 dcterms 元素中的关联关系 hasPart,用以表示智库产品(prov:asset)包含智库报告(fabio: Report)、智库论文(bibo: Article)、博客(blog)、评论(commentary)、智库政策(policy)。

在设计智库资源元数据元素集时,主要以 DC 元数据标准构建智库资源的核心元数据元素集合,同时融合专门元数据元素对核心元数据进行补充和扩展,最终形成适用于智库资源的集成元数据标准。在完成集成元数据标准建设后,在智库资源实体化的基础上对其属性以及相互关系进行定义,并按照集成元数据标准建立智库资源的“实体-属性-关系”网络图(见图 4),为构建智库资源的集成元数据本体奠定基础。

集成元数据本体解决了智库资源元数据语义异构的问题,包括资源采用不同元数据方案所造成的微观结构的异构问题以及资源对象之间存在的复杂的关联关系,使智库资源有了基本的微观结构。但智库资源包含了不同类型多源异构资源,仅构建智库资源的集成元数据本体不足以实现全球智库信息集成服



