

工程科技知识组织系统建设规范研究报告

中国科学技术信息研究所

2016-12-22

目录

1 引言	1
2 知识组织系统建设意义.....	1
2.1 知识组织及知识组织系统概念界定.....	1
2.2 知识组织系统建设方法.....	2
3 知识组织系统具体实施.....	3
3.1 核心词汇选择.....	3
3.2 知识组织系统建设指导规范.....	4
4 知识组织系统建议内容结构.....	7
5 知识组织系统建设建议流程.....	9
参考文献.....	14

1 引言

本规范研究报告主要内容包括：界定本项目中的知识组织及知识组织系统的概念，描述知识组织系统建设的方法与途径，阐明知识组织系统建设原则和具体实施目标。针对知识组织系统建设任务制定指导规范，给出知识组织系统建设的内容、结构、建议流程及其建议周期。最后，明确了系统建设机构及其相应职责、列出了相关参考文献。

2 知识组织系统建设意义

知识组织系统建设是中国工程科技知识中心建设和服务过程中的一项重要任务，也是支撑知识服务的重要基础设施。通过知识组织系统建设能够更好地组织中国工程科技知识中心的数据资源，进而为政府机构、院士及行业专家团队、研究院所等科研机构、大学、企业等各类用户以及国家重大科技战略实施提供优质的知识服务。

工程领域知识组织系统有利于实现高效、深入的信息服务和知识服务。通过工程知识组织系统的持续建设，能够建立领域性的知识框架，优化数据资源配置，提高知识的加工效率和一致性，增强知识导航和知识服务的效果。

2.1 知识组织及知识组织系统概念界定

“知识组织”是以知识为对象的诸如整理、加工、表示、控制等一系列组织化过程及其方法。“知识组织系统”是服务于知识组织的术语和符号体系。知识组织系统包含的范围较广，分类法¹和主题法是知识组织两大经典方法，而叙词表、词系统和本体²等综合反应了分类法与主题法的思想精髓，是目前知识组织

1分类法一般指根据类目之间关系组织起来，并配有一定标记符号的类分信息资源的工具。本规范中不强调类似中国图书分类法和国家专利分类法这样的分类法，而是专注到英国 BS-8723 标准提及，道琼斯公司力推以及在知识管理等应用中更为实用的一种规模小，层级少，无标记符号的知识分类表，其形式与叙词表类似，组织的内容是类目，但是也通过术语词汇或短语来表达。

2本体也称实用分类法是共享概念模型的形式化规范说明，给出构成相关领域词汇的基本术语和关系，以及利用这些术语和关系构成的规定这些词汇外延的规则的定义，并建立约束规则和推理规则。

系统的主流表现。

针对叙词表的知识内容表现不够、本体又无法真正实现的现状，本规范的知识组织系统基本以词系统为基础。其主要原因：1) 词系统向下可以兼容传统的术语词典、叙词表等，向上可以视为轻量级的本体，具备一定的词表和本体的特色。2) 尽管词系统不是最先进的理论研究产物，但词系统是当前知识组织系统的较好实践。词系统可以基于新的研究成果，不断地进行丰富与扩展，比如向真正的本体方向转化与进化。

本规范中的知识组织系统是一种定义、组织和表述领域知识的术语和符号的体系。它继承了叙词表主要知识架构，并进行了适当的知识扩充，提供特定领域的语义资源基础。在实际应用中，它可以作为一种语义工具，可用于标引和检索、科技分析和监测、信息抽取和服务、知识导航中。

2.2 知识组织系统建设方法

中国工程科技知识组织系统分为总中心知识组织系统和分中心知识组织系统，总中心知识组织系统在收集整理各类工程科技知识组织系统资源和分中心提供的知识组织系统基础之上，进行整合和集中服务。分中心知识组织系统在自有知识组织系统基础、数据资源和可获取的其他知识组织系统之上，建设修订和完善。

为了兼容不同的类型的知识组织系统建设和统一管理的需求，中国工程科技知识组织系统以词系统为基础平台进行知识组织系统建设，结果统一输出为 OWL 格式，根据知识组织系统本身的不同，设置相应的输入/输出模板。在术语的选择和词间关系上以叙词表等同关系、属分关系和相关关系三大类为基础，必要时可以扩展，例如：是否需要将关系进一步细化，细化为多少以及是否选用定义、属性、分类映射等可以根据实际情况选择。

针对各单位知识组织系统情况，建设方法如下表所示：

基础资源情况	处理策略	处理工具
有知识组织系统	知识检验、逻辑检验、知识丰富	语料库工具、逻辑校验工具、关系发现工具
有相关的同语言的知识组织系统	知识组织系统切分、知识组织系统融合	社区发现工具、逻辑校验工具、概念加工平台
有相关的不同语言的知识组织系统	汉化翻译、知识调整	机器翻译工具、语料库工具、逻辑校验工具
有相关语料资源	统计选词、关系发现	语料库工具、关系发现工具

3 知识组织系统具体实施

3.1 核心词汇选择

依照相关工程科技学科领域汉语词类的特点，选择合适的词性和词汇类型。词汇的选择以领域名词（包括专有名词）、动名词为主，选择的词型包括词汇和词组。

在选词的内容范围上，要求核心词条能够覆盖领域的大部分数据资源，以及相关领域涉及的多学科、业务领域词汇。

词汇知识结构设计

系统的知识结构可以兼容叙词表和本体。词汇知识结构包括：词汇定义、词汇间关系、词汇分类信息以及词汇属性等。

（1）建设核心词的定义知识，简单并明确地描述核心词的涵义。

（2）建设与核心词有关的三类关系，包括同义关系，属分关系和相关关系。通过同义关系的建立，便于计算机处理与检索同一主题的内容，同时建立同义词知识库；相关关系的建立便于计算机扩大该主题的检索途径，以及为主观题判卷提供更为全面的领域词汇的语义支撑；属分关系的建立，能够为标引，如按族检索提供层级关联词汇的支持。在实际建设的过程中，这三类词间关系需要根据实际的领域知识做进一步的细化。

（3）通过词汇与分类映射的建立，能够为工程院知识内容按分类浏览提供支持。

（4）通过词汇属性建设，能够为精确描述词汇以及向本体过渡提供支持。