

中国科学技术信息研究所 多语言文献知识库构建

中国科学技术信息研究所

2020年12月10

目 录

1 引言.....	6
2 文献资源的管理和组织的发展.....	6
2.1 传统图书资料组织管理.....	6
2.2 电子文档组织管理.....	6
2.3 网络化数字资源组织管理.....	7
2.4 语义 web 资源组织管理.....	7
3 知识库的构建方法.....	7
3.1 知识表示方法.....	7
3.1.1 框架.....	7
3.1.2 脚本.....	8
3.1.3 语义网络.....	8
3.1.4 逻辑.....	9
3.2 典型的知识库系统.....	9
3.2.1 Cyc 常识知识库.....	9
3.2.2 WordNet 英文概念知识库.....	10
3.2.3 HowNet 中文概念知识库.....	10
3.2.4 典型知识库系统的总体特点.....	12
3.3 利用语义 web 技术构建的知识库.....	11
3.3.1 Wikidata.....	11
3.3.2 Freebase.....	11
3.3.3 Yago.....	12
3.3.4 Linked Data.....	12
3.3.5 国内开发的知识库系统.....	12
3.3.6 利用语义 web 技术构建知识库的总体特点.....	13
4 多语言文献知识系统的构建.....	13
4.1 多语言文献知识模型.....	13
4.2 本体知识库的组织与表示.....	14
4.2.1 RDF 资源描述框架.....	16
4.2.2 OWL 网络本体语言.....	18
4.2.3 知识三元组序列化方法.....	24
4.3 本体知识模型构建工具介绍.....	25
4.4 知识三元组的查询模式.....	34
4.4.1 基本查询.....	37
4.4.2 数据集合查询.....	41
4.4.3 功能性查询.....	44

4.5 多语言文献知识三元组的生成方法.....	46
4.5.1 三元组的生成过程.....	46
4.5.2 实例唯一标识的生成.....	47
4.5.3 知识资源模板.....	48
4.6 知识资源的管理和发布.....	53
4.6.1 关系数据转换工具 D2RServer.....	54
4.6.2 高性能三元组存储和检索工具 Virtuoso.....	56
4.6.3 三元组发布和浏览工具 Pubby.....	57
4.6.4 轻量级三元组存储和管理工具 RDF4J.....	58
5 多语言文献知识系统的管理和维护.....	59
5.1 服务接口的实现方法.....	59
5.1.1 SOAP 式 Web 服务.....	59
5.1.2 RESTful 式 Web 服务.....	61
5.2 系统目录结构.....	67
5.3 管理界面.....	68
5.4 服务接口的调用.....	71
5.4.1 知识服务接口的启动和调用.....	71
5.4.2 客户端界面.....	73
5.4.3 数据检索模式.....	74
6 多语言文献知识内容的展示.....	79
6.1 科研要素关联图.....	79
6.2 科研要素统计图.....	81
7 研究总结.....	83
7.1 知识规模和性能统计.....	83
7.2 技术特点.....	85

插图清单

图 4-1	科研要素知识模型.....	14
图 4-2	OWL1 的子语言.....	19
图 4-3	ActiveOntology 面板.....	26
图 4-4	Classes 面板.....	27
图 4-5	ObjectProperties 面板.....	28
图 4-6	DataProperties 面板.....	29
图 4-7	AnnotationProperties 面板.....	30
图 4-8	Individuals 面板.....	31
图 4-9	OWLViz 面板视图配置安装路径-1.....	32
图 4-10	OWLViz 面板视图配置安装路径-2.....	32
图 4-11	OWLViz 面板视图配置安装路径-3.....	33
图 4-12	SPARQLQuery 面板.....	34
图 4-13	D2R 平台的功能结构.....	55
图 4-14	D2R 平台的服务启动.....	55
图 4-15	OpenVirtuoso 的功能结构.....	56
图 4-16	Pubby 的功能结构.....	57
图 4-17	轻量级三元组存储和管理工具 RDF4J.....	58
图 5-1	多语言文献知识库系统的目录结构.....	67
图 5-2	多语言文献知识库的管理界面的基本布局-1.....	68
图 5-3	多语言文献知识库的管理界面的基本布局-2.....	69
图 5-4	多语言文献知识库的管理界面的基本布局-3.....	69
图 5-5	多语言文献知识库的管理界面的基本布局-4.....	70
图 5-6	知识服务接口的启动和调用.....	71
图 5-7	多语言文献知识库系统客户端界面 SPARQL 检索框.....	73
图 5-8	多语言文献知识库系统客户端界面 SPARQL 检索结果显示区域.....	74
图 6-1	科研要素关联图-1.....	79
图 6-2	科研要素关联图-2.....	80
图 6-3	科研要素关联图-3.....	80
图 6-4	科研要素统计图-1.....	81
图 6-5	科研要素统计图-2.....	82
图 6-6	科研要素统计图-3.....	82
图 7-1	多语言文献科研要素本体模型的层级结构.....	84
图 7-2	多语言文献知识库数据入库运行过程.....	85