

科技知识组织体系版本管理系统设计与实践应用*

刘丹军¹ 付鸿鹄² 文奕¹ 胡正银^{1,3} 杨宁¹ 向彬¹ 钱力^{2,3} 刘春江¹

¹(中国科学院成都文献情报中心 成都 610041)

²(中国科学院文献情报中心 北京 100190)

³(中国科学院大学 北京 100049)

摘要:

[目的] 设计针对科技知识组织体系(STKOS)数据进行多版本及版本内修订管理的方案,并研建实际服务系统。

[应用背景] STKOS 共享服务平台要求实现对不同版本数据的管理、发布及应用,并及时发布当前版本的修订信息。

[方法]

在定义历史版本、当前正式服务版本、临时版本三种 STKOS 版本类型的基础上,设计 STKOS 版本变更信息数据结构与 STKOS 多版本管理、版本内修订管理流程。然后,基于医药卫生范畴 STKOS 数据,研建 STKOS 版本管理系统。

[结果] 在千万量级数据场景下,构建了 STKOS 版本管理系统,实现了 STKOS 多版本管理与版本内修订管理。

[结论] 本系统可同时支持海量数据 STKOS 多版本管理及版本内数据修订管理。

关键词: 科技知识组织体系 版本管理 大数据 知识服务

分类号: G353.1

Study on STKOS Version Management

Liu Danjun¹ Fu Honghu² Wen Yi¹ Hu Zhengyin^{1,3} Yang Ning¹ Xiang Bin¹ Qian Li^{2,3}

Liu Chunjiang¹

¹(Chengdu Document and Information Center, Chinese Academy of Sciences, Chengdu 610041,China)

²(National Science Library, Chinese Academy of Sciences, Beijing 100190,China)

³(University of Chinese Academy of Sciences, Beijing 100049,China)

Abstract:

[Objective]

This paper focuses on designing a framework for STKOS multi-versions and inter-version revise management and implementing a STKOS version management system based on the framework.

[Context] The sharing service platform of STKOS defines the content of STKOS version management, including multi-versions and inter-version revise management.

[Methods]

Firstly, we define three types of versions of STKOS, namely historic version, active service version, temporary version. Then we design the data structure of STKOS changing information and a framework of STKOS multi-versions and inter-version revise managements. Finally, we implement a STKOS version management system based on the framework and the medical STKOS data.**[Results]** In ten millions data size, this system implements the version management of STKOS.

[Conclusions] The system can support multi-versions and inter-version revise management of large scale STKOS data at the same time.

* 本文系国家科技支撑计划项目“科技知识组织体系共享服务平台建设”子课题“科技知识组织体系(STKOS)发布服务系统”(项目编号:2011BAH10B03-2)的研究成果之一。

Keywords: STKOS Version management Big data Knowledge service

1 引言

为了更有效利用海量科技文献信息, 国家科技图书文献中心牵头组织实施了国家科技支撑计划“面向外文科技文献信息的知识组织体系建设与应用示范”项目^[1]。科技知识组织体系(Science and Technology Knowledge Organization System, STKOS)发布服务系统是该项目面向用户的服务系统, 基于 STKOS 实现了知识组织体系、知识本体、科研本体的浏览、查询、推理、统计、定制、整合、嵌入等服务, 并提供多种方式浏览和利用 STKOS 及 STKOS 的子集, 以满足不同机构、不同用户对领域知识本体的需求, 达到最大程度地共享和重用科技知识组织体系成果的目的^[2]。

版本管理已广泛应用于软件开发、文档管理中。科技知识组织体系版本管理对各种不同类型版本的 STKOS 进行导入、浏览、存档、导出等管理, 为用户跨版本利用 STKOS 数据提供支持。通过 STKOS 版本管理, 可以回溯不同时期、不同版本 STKOS 的变迁; 也可及时反映当前 STKOS 版本内的修改、变动情况。STKOS 版本管理系统为 STKOS 发布服务系统各个功能模块提供跨版本的数据接口, 是整个 STKOS 服务系统的重要组成部分。

本文在深入分析、比较现有知识组织体系版本管理的特点、适用范围基础上, 结合 STKOS 发布服务系统的具体功能需求, 顺应未来 STKOS 版本数据可能的发展趋势, 设计了 STKOS 版本管理方案。该方案可同时对不同时期版本的 STKOS 数据及当前服务 STKOS 版本内部的修订数据进行管理, 支持对 STKOS 各历史版本及当前服务版本数据的跨版本应用, 为未来正式的 STKOS 多版本发布、综合应用奠定基础。

2 需求及技术思路

2.1 知识组织体系版本管理调研

知识组织体系的版本管理可分为两种: 多版本管理与版本内修订管理。前者对多个正式发布知识组织体系版本的变更信息进行管理, 让用户可跨版本应用相关数据; 后者通过记录当前正式服务版本内修订信息, 让用户可及时使用到知识组织体系的最新数据, 而不必等到最终的正式版本发布。对美国国家医学图书馆的统一医学语言系统(Unified Medical Language System, UMLS)^[3]、联合国粮农组织研建的基于简单知识组织系统(Simple Knowledge Organization System, SKOS)的词表管理系统(Vocabulary Bench, VocBench)^[4]及化学文摘(Chemistry Abstract, CA)光盘数据库版本管理^[5]进行调研, 结果如表 1 所示:

表 1 UMLS、VocBench 与 CA 光盘数据库版本管理比较

名称	多版本管理	版本内修订管理	版本数据
UMLS	通过变更文件来记录版本间元素变更信息。只记录发生变更的元素 ID, 如 CUI、AUI 等; 及变更类型, 如元素新增、概念合并等	不对外发布正式版本的内部修订管理信息	正式版本数据是一系列小文件; 版本变更信息记录在 ChangeLog 文件中

VocBench	每一个版本单独保存为一个数据库；不记录不同版本间变更信息	通过数据库 OntoChange 表记录版本内元素修订情况，如名称变动、关系变动等	正式版本数据是一个单独 SKOS 文件
CA 光盘数据库	提供年度存档光盘文件，作为当年历史版本	按月提供更新到当月的最新年度数据，作为本年度最新数据	年度、月份更新数据都直接以数据库文件形式存放

从表 1 可看出，UMLS 通过历史版本变更数据文件记录多个 UMLS 正式版本之间的变更信息，只能反映 UMLS 不同版本间变更信息，而不能及时反映当前 UMLS 版本变更信息；VocBench 通过变更信息表记录可及时反映同一版本内词表修订情况，但是不能反映跨版本变更信息，只适用于同一版本数据的加工管理；CA 光盘数据库通过按月更新来及时反映本年度数据更新情况，每次版本更新，都会完全覆盖上一版本的数据，无法提供跨版本检索与查新服务。

2.2 STKOS 版本类型设计

STKOS 版本管理包含：STKOS 多版本管理与 STKOS 版本内修订管理两部分。STKOS 多版本管理对正式发布的多个 STKOS 版本数据及其变更信息进行管理，可提供历史版本查询及跨版本查询服务；而 STKOS 版本内修订管理则通过对当前服务 STKOS 版本的小变更信息进行管理，便于用户及时使用 STKOS 最新数据。与表 1 调研的三个典型知识组织体系版本管理系统相比，同时考虑到 STKOS 正式版本发布周期一般较长，STKOS 版本管理系统的优点是：既能通过多版本管理反映不同正式 STKOS 版本间变更信息，又能通过版本内修订管理让用户及时了解当前正式服务版本的最新信息。

(1) STKOS 多版本管理：管理对象是不同时期发布的正式版本 STKOS。其管理内容包括：

- ①确定管理维护周期，及时发布最新版本，以确保检索默认项为最新版本的内容；
- ②完整保留历史版本的信息，同时通过版本变更信息表记录每一个版本中的变更信息；
- ③STKOS 多版本数据可视化对照显示。

(2) STKOS 版本内修订管理：管理对象是当前正式服务版本 STKOS。其管理内容包括：

- ①根据最新加工数据，对 STKOS 当前正式服务版本内容进行即时更新；
- ②通过修订变更信息表记录每一次修订的变更信息。

为满足上述需求，本文定义了三种 STKOS 版本类型：历史版本、当前正式服务版本、临时版本。

(1) 历史版本属于长期保存存档性质，不再接受任何修改，主要提供历史版本查询、对照浏览、跨版本查询等功能。

(2) 当前正式服务版本是在最近一期历史版本的基础上，整合了 STKOS 版本内修订数据。与最近一期历史版本相比，它能较及时反映在下一个正式版本发布之前，可能

出现的一些较小内容变更，为其他应用提供数据支持。

(3)临时版本负责同步接收来自加工平台 STKOS 最新修订数据，并进行临时存储。当后台词表管理员认为当前正式服务版本可接受该修订变化后，即根据临时版本的最新修订数据，更新当前正式服务版本数据。

历史版本有多个，而当前正式服务版本与临时版本只有一个。在数据存储方面，历史版本、当前正式服务版本、临时版本数据结构基本一致，存储在一套表中。当前正式服务版本及历史版本之间的数据变更信息通过变更信息表进行存储。变更信息表只记录 STKOS 元素唯一标识符(Unique Identifier, UI)针对某一版本发生何种变更，而不记录变更前后的详细信息。由于 UI 是 STKOS 元素跨版本的唯一标识符，可通过元素 UI 在不同版本间查询该元素的详细变更信息。当前正式服务版本内的修订信息则完整记录在临时版本中。三种 STKOS 版本类型内容如表 2 所示：

表 2 三种 STKOS 版本类型内容

版本类型	版本性质	版本来源	版本功能
历史版本	长期保存存档版本；不再接受任何修订	由当前正式服务版本归档生成	主要提供历史版本查询、跨版本查询等
当前正式服务版本	活动版本；通过修订管理接受版本内小修订	由最新 STKOS 版本数据导入生成	为知识引擎、发布服务、第三方系统等提供数据
临时版本	保存当前正式服务版本的修订信息；发布到当前正式服务版本后即清空	同步加工平台 STKOS 最新修订数据生成	为当前正式服务版本提供修订数据临时存储空间，用于词表管理员对 STKOS 修订发布前进行审核

2.3 STKOS 版本管理系统框架及流程设计

STKOS 版本管理系统在整个科技知识组织体系发布服务系统中处于承上启下的位置，接受来自 STKOS 加工平台的原始 STKOS 版本文件作为输入，以 Solr 索引的方式向 STKOS 发布服务系统其他模块及第三方应用提供数据接口。Solr 是一个高性能、企业级搜索应用服务器，对外以 XML 文件形式提供类似于 Web Service 的 API 接口，被广泛应用于搜索中^[6]。STKOS 版本管理系统在科技知识组织体系发布服务系统中位置如图 1 所示：

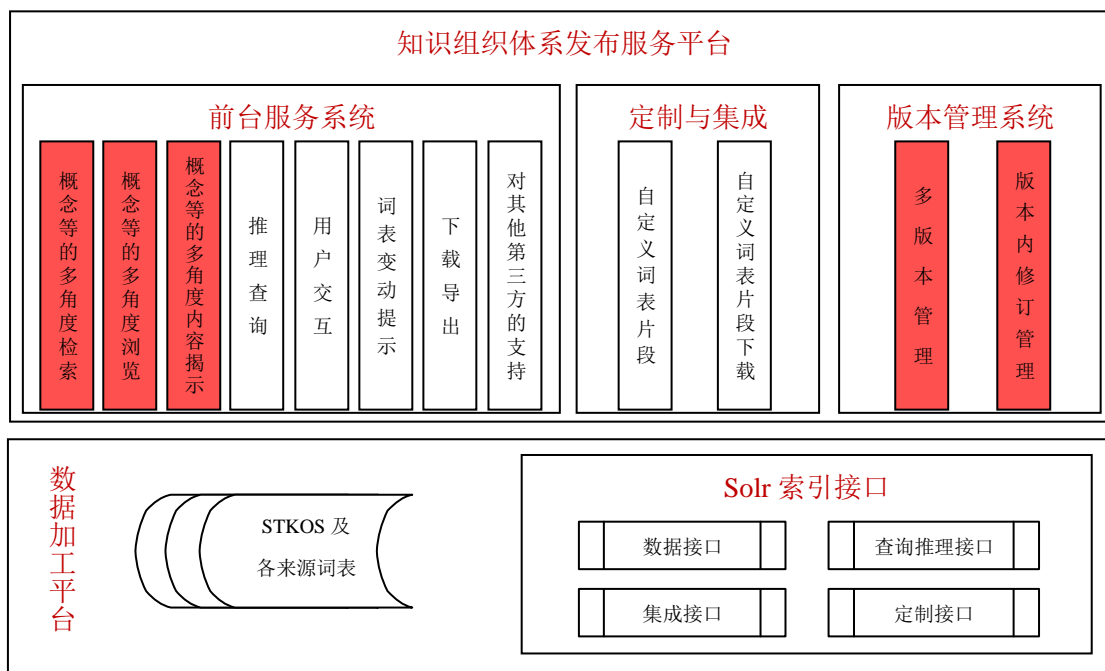


图 1 科技知识组织体系发布服务系统总体框架

STKOS 多版本管理流程包括：STKOS 最新版本数据导入、STKOS 完整性校验、STKOS 当前正式服务版本归档、STKOS 最新版本数据审核发布、重建 STKOS 当前正式服务版本索引，具体流程如图 2 所示：

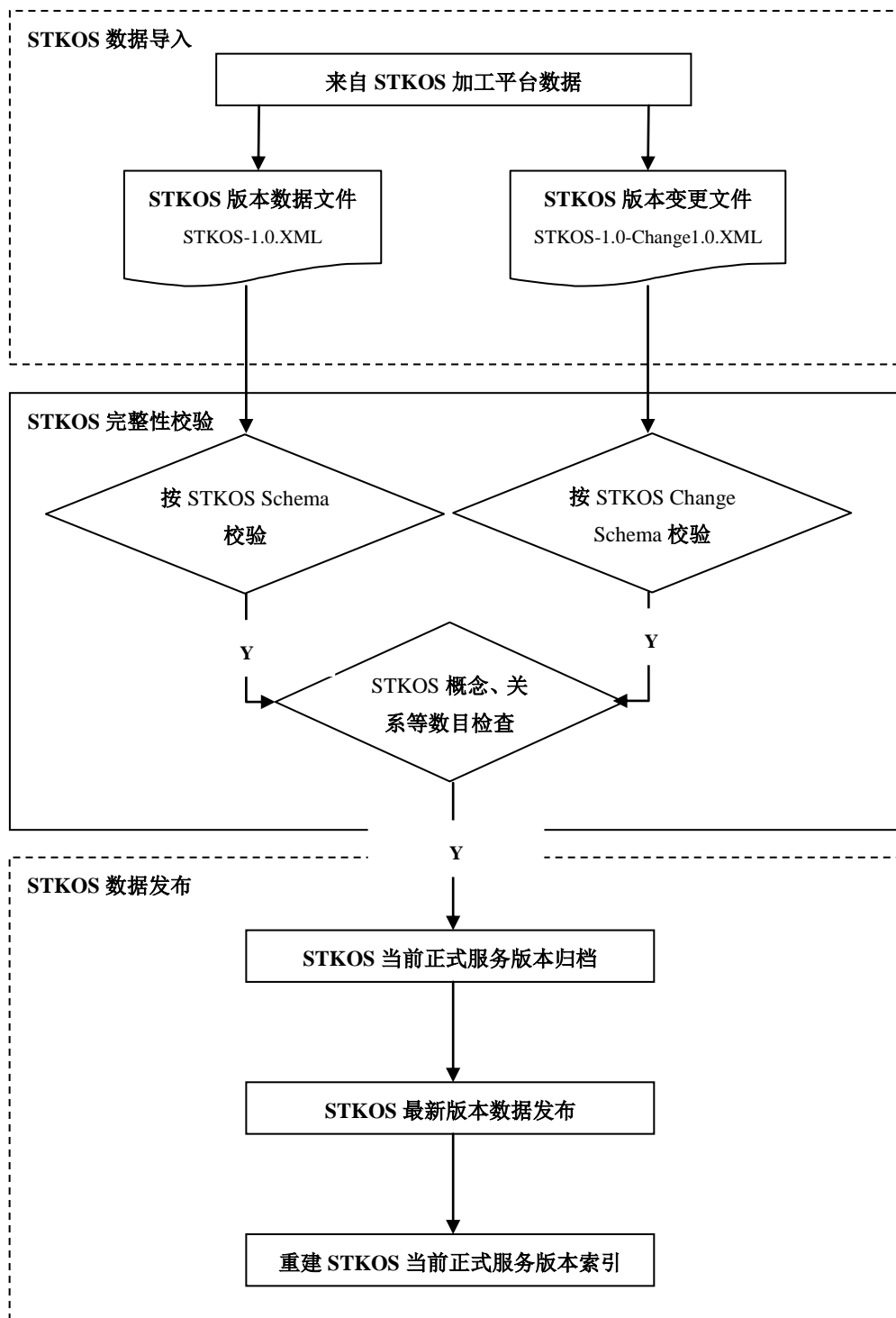


图 2 STKOS 多版本管理流程图

STKOS 最新版本数据导入之后，要进行完整性校验。STKOS 数据文件通过完整性校验之后，需要对当前正式服务版本进行归档操作，将其变更为历史版本。然后再将导入的最新版本 STKOS 数据作为当前正式服务版本发布；最后重建 STKOS 当前正式服务版本 Solr 索引，供前台发布系统或第三方服务系统调用。

STKOS 版本内修订管理流程与多版本管理类似，包括以下 4 步：STKOS 最新修订

数据导入、修订数据完整性校验、更新当前正式服务版本修订数据、重建 STKOS 当前正式服务版本 Solr 索引。

2.4 STKOS 版本管理系统数据结构设计

版本管理系统采用关系型数据库 MySQL 来存储 STKOS 词表本身及其版本变更数据,数据结构设计包括 STKOS 词表数据结构设计与版本变更信息数据结构设计两部分。

(1) STKOS 词表数据结构设计

STKOS 词表数据结构设计参考 ISO 25964-1 数据模型、STKOS 发布元数据、UMLS 数据模型。ISO 25964-1 数据模型定义了叙词表相关标准规范,提供了数据模型建议^[7]; STKOS 发布元数据提供了 STKOS 的数据模型^[8]; UMLS 则是成熟、应用广泛的词表。

STKOS 词表数据结构参考 ISO 25964-1 数据模型,遵循 STKOS 发布元数据规范,允许数据的适当冗余。STKOS 发布元数据规范中定义了来源术语、基础术语、STKOS 规范概念、来源词表、范畴表与范畴类 6 种数据类型。根据对 STKOS 发布元数据的具体分析,术语与概念关系主要建立在概念和基础术语之上。因此,词表数据以基础术语、概念、来源词表、范畴为主。基于上述设计思路,版本管理系统共设计了基础术语表(保存基础术语数据)、概念表(保存 STKOS 规范概念数据)、概念定义表(保存概念所有的定义)、来源词表(保存来源词表相关数据)、范畴表(保存 STKOS 范畴数据)、范畴关系表(保存 STKOS 范畴之间的关系)、概念范畴表(保存 STKOS 规范概念与范畴之间的关系)、关联关系表(保存概念与术语之间的关联关系)、层级关系表(保存关系的层级信息)等 9 个表,用于管理 STKOS 版本数据。

(2) STKOS 版本变更信息数据结构设计

STKOS 版本变更信息数据结构设计参考了 UMLS 版本变更文件格式^[3]、词表管理平台 VocBench 中词表最新变更表^[4]。UMLS 版本变更文件记录 UMLS 正式版本针对前一版本间的重大变更信息;VocBench 通过词表变更表来记录同一版本中词表修订情况。两者结合起来,正好可满足 STKOS 版本管理系统既要保存正式发布的多版本 STKOS 数据,又要及时反映当前 STKOS 最新版本的变更信息的要求。

版本管理系统共设计了版本信息表(保存 STKOS 版本数据)、单元素变更信息表(保存术语或概念单元素变更信息)、二元关系变更信息表(保存术语与概念之间二元关系变更信息)、概念合并表(保存概念跨版本合并变更信息)、术语移动表(保存术语跨版本移动信息)、版本内修订信息表(保存版本内数据修订变更信息)、概念停用表(保存 STKOS 概念停用信息)、术语停用表(保存术语停用信息)等 8 个表,用于管理 STKOS 版本变更信息数据。

3 具体实现方案

3.1 STKOS 多版本管理模块

STKOS 多版本管理模块主要包括 STKOS 数据导入、STKOS 完整性校验、STKOS 数据发布及 STKOS 版本变更对照 4 个功能。

(1) STKOS 数据导入

STKOS 数据导入是一个与 STKOS 加工平台数据交换接口。当有新版本数据需要导入时,加工平台提供符合 STKOS Schema 规范的正式版本数据文件及符合变更信息

STKOS Change Schema 规范的版本变更信息文件。STKOS 数据导入模块提供大数据 Web 上传导入功能。以医药卫生范畴 STKOS 词表为例，该词表的数据文件包括：描述词表中包含术语的文件 termRecords.xml、词表中学科范畴的文件 stCategoryRecords.xml、描述概念间关系的文件 accociativeRelsRecords.xml 等共 11 个数据文件。上传 STKOS 版本数据文件界面如图 3 所示：

上传XML文件

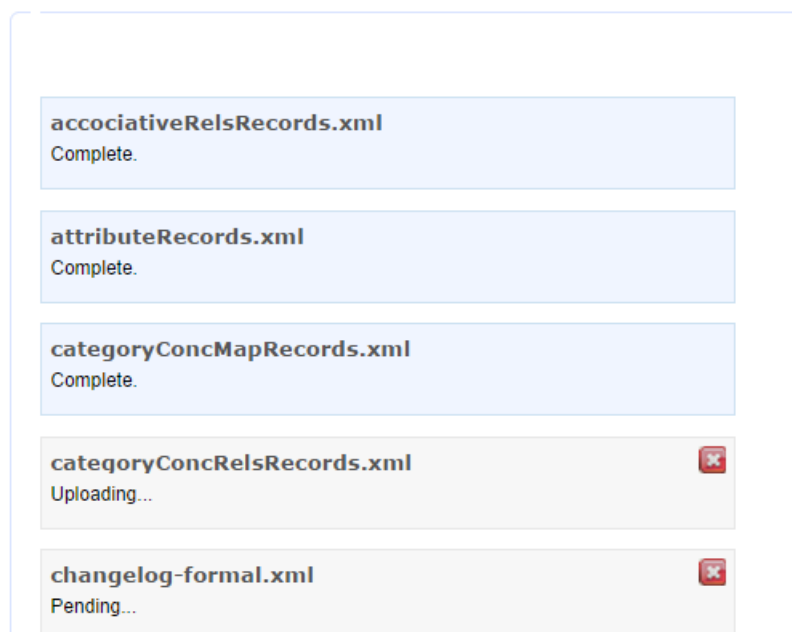


图 3 STKOS 版本数据上传界面

(2) STKOS 完整性校验

STKOS 数据量很大，如医药卫生范畴 STKOS 词表中记录属性的 attributeRecords.xml 文件约 10GB，数据量超过 1 000 万条。为保证数据完整性，必须在正式发布前，对导入的 STKOS 数据进行完整性校验。完整性校验包含两个层面。首先，利用 STKOS 数据的 XML Schema 进行语法层面上的校验，确保导入的 STKOS 文件中各元素及属性均符合预定义的语法规范。其次，对 STKOS 文件的各元素，如概念、关系等的数目进行校验，确保 STKOS 文件的数据完整。导入之后，系统自动进行完整性校验，校验结果如图 4 所示：

文件类型	XML文件名	XML Schema	文档记录数	实际记录数	验证结果	详细信息
日志文件	changelog-formal.xml	changelog-formal.xsd	1342	1342	正确	[耗时:0.131秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:1342,1342.
文件类型(序号)	XML文件名	XML Schema	文档记录数	实际记录数	验证结果	详细信息
数据文件1	termRecords.xml	termRecords.xsd	919701	919701	正确	[耗时:1352.161秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:919701,919701.
数据文件2	thesaurusRecords.xml	thesaurusRecords.xsd	54	54	正确	[耗时:0.138秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:54,54.
数据文件3	stTranslationRecords.xml	stTranslationRecords.xsd	478	478	正确	[耗时:0.761秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:478,478.
数据文件4	stDefinitionRecord.xml	stDefinitionRecord.xsd	65300	65300	正确	[耗时:70.276秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:65300,65300.
数据文件5	stCategoryRecords.xml	stCategoryRecords.xsd	689249	689249	正确	[耗时:944.055秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:689249,689249.
数据文件6	hierarchyRecords.xml	hierarchyRecords.xsd	66	66	正确	[耗时:0.062秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:66,66.
数据文件7	categoryConcMapRecords.xml	categoryConcMapRecords.xsd	1499173	1499173	正确	[耗时:8611.055秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:1499173,1499173.
数据文件8	categoryConcReIsRecords.xml	categoryConcReIsRecords.xsd	125	125	正确	[耗时:0.109秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:125,125.
数据文件9	attributeRecords.xml	attributeRecords.xsd	14857738	14857738	正确	[耗时:22104.58秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:14857738,14857738.
数据文件10	accocativeReIsRecords.xml	accocativeReIsRecords.xsd	7939024	7939024	正确	[耗时:12574.11秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:7939024,7939024.
数据文件11	definitionRecords.xml	definitionRecords.xsd	121398	121398	正确	[耗时:140.743秒]1)XML较验合法. 2)XSD较验合法. 3)记录数[文档,实际]:121398,121398.

图 4 STKOS 数据完整性校验结果

(3) STKOS 数据发布

STKOS 数据发布包括：当前正式服务版本归档、导入数据发布为当前正式服务版本、重建当前正式服务版本 Solr 索引三个步骤。前两个步骤主要是改写版本状态信息，即将前一个 STKOS 从当前正式服务版本改为历史存档版本；并将新导入 STKOS 数据标记为当前正式服务版本。当前正式服务版本的 Solr 索引是版本管理系统对外提供数据服务的接口；当版本数据变更后，必须重建 Solr 索引。考虑到重建 Solr 索引数据量很大，耗时较长，系统采用按照业务分段方式，即每次取一定数量（如 2 万条记录）在后台增量重建 Solr 索引。

在 STKOS 版本列表显示每个版本的概况，包括：来源词表数、覆盖的学科范畴(范畴)、来源术语总数、概念及关系总数、与前一版本相比较变化的概况(如概念间关系变化情况)以及版本活动状态等信息。版本列表默认按版本发布时间倒序排列。STKOS 数据发布操作界面如图 5 所示：

所在位置：版本管理(汇总表)									
版本	来源词表	范畴	来源术语	概念	关系	版本状态	发布时间	操作	变更汇总
STKOS2013	5	131	919608	652800	4781020	当前	2013-05-08	归档为历史版本 发布到知识库 重建Solr索引 版本变更对照	合并 :12 移动 :12 关系 :18 [(关系) 新增 :6, 修改 :6, 删除 :6] 单元素 :18 [(单元素) 新增 :6, 修改 :6, 删除 :6]

图 5 STKOS 数据发布操作界面

对于每一个版本 STKOS 数据，系统还可详细显示来源词表、学科范畴、来源术语、概念、关系等信息。学科范畴详细信息界面如图 6 所示：

所在位置：查询范畴词表

查询结果：共 1331 条记录，当前是第 1 页，共 14 页 首页 上一页 1 2 3 4 5 6 7 8 9 10 下一页 尾页

序号	范畴词表
1	DNA病毒感染
2	Q热
3	RNA病毒感染
4	休耕制
5	回肠炎
6	破裂, 冻疮, 冻伤
7	线虫病(线虫感染)
8	肺坏疽

图 6 STKOS 学科范畴详细信息列表界面

(4) STKOS 版本变更对照

STKOS 版本变更对照功能可揭示不同版本及当前正式服务版本之间的最新变更信息。新版本的变动信息可通过列表方式查看，或通过版本间的对照直观显示。通过版本对照视图，用户可以选择两个版本的 STKOS 数据进行对比，界面如图 7 所示：

所在位置：版本变更对照

变更概要信息：
 合并 :12
 移动 :12
 关系 :18 [新增 :6, 修改 :6, 删除 :6]
 单元素 :18 [新增 :6, 修改 :6, 删除 :6]

当前版本(仅显示变更数据)

- 合并 (12)
 - mrgcui:1
 - cui1:C0000612
 - mrgcui:2
 - cui1:C0000618
 - mrgcui:3
 - cui1:C0000644
 - mrgcui:4
 - cui1:C0000657
 - mrgcui:5
 - cui1:C0000661
 - mrgcui:6
 - cui1:C0000666

对照版本(仅显示变更数据)

- 合并 (12)
 - mrgcui:1
 - cui1:C0000612
 - cui2:C0000617
 - mrgcui:2
 - cui1:C0000618
 - cui2:C0000641
 - mrgcui:3
 - cui1:C0000644
 - cui2:C0000656
 - mrgcui:4
 - cui1:C0000657
 - cui2:C0000658

图 7 STKOS 版本变更对照图

系统显示两个版本 STKOS 词表间的变更概要信息及具体变更情况，并列两个版本词表树形结构。对于发生变更的元素，可以根据变更类型（如概念合并、关系变化等）查看变更的详细信息；通过元素 ID，可查看不同版本中该元素的具体信息。

3.2 STKOS 版本内修订管理模块

STKOS 版本内修订管理模块包括：修订数据导入、修订数据完整性校验、修订数据发布三步。

STKOS 修订数据是 STKOS 正式版本数据的一个子集，仅包含最新修订数据信息，没有变更信息文件。STKOS 修订数据导入之后存储在 STKOS 临时版本数据库中。通过数据完整性检查与 STKOS 词表管理员审核之后，即可予以发布。修订数据发布采用更新当前正式服务版本 STKOS 数据方式进行，同时采用增量更新的方式重建正式服务版本 Solr 索引。版本内修订管理界面如图 8 所示：

序号	概念或术语标识1	概念或术语标识2	标识类型1	标识类型2	术语标识1	术语标识2	旧关系标识	新关系标识	变更活动	变更时间	操作
1	C0022672	C0027721			A0012490	A0012490			移动	2013-05-09 17:34:09.0	发布 暂缓发布
2	C0039437	C0018520			A0011395	A0011395			移动	2013-05-09 17:34:09.0	发布 暂缓发布
3	C0000612	C0000617							合并	2013-05-09 17:24:13.0	发布 暂缓发布
4	C0000657	C0000658							合并	2013-05-09 17:24:13.0	发布 暂缓发布
5	A0001895	A1389616	AUI	AUI			R00689636	R00689636	修改	2013-05-09 16:21:03.0	发布 暂缓发布

图 8 STKOS 版本内修订管理界面

4 应用效果评估

目前，STKOS版本管理系统已用于科技知识组织体系发布服务系统后台管理中，并对医药卫生范畴STKOS数据进行管理，总体状况良好。系统具体性能指标如下：

- (1) 导入速度：医药卫生范畴 STKOS 数据共 11 个 XML 文件，文件大小约 25GB，总数据量约 3 000 万条。可在 20 小时内对这些数据文件进行完整性校验、数据库导入、发布及索引。
- (2) 响应时间：可在毫秒级内响应 STKOS 版本数据的浏览、检索、版本变更数据对照展示等请求。
- (3) 并发性能：在 STKOS 数据导入时，采用独占模式；在 STKOS 数据管理时，支持 100 个以上并发请求。

从科技知识组织体系发布服务未来发展趋势看，系统还存在需要完善的地方，建议从以下三方面改进：

- (1) STKOS 完整性校验、导入性能有待提高：完整 STKOS 数据规模会更大，按照现有技术框架，完整性校验与导入花费时间会更长。未来，一方面可考虑与加工平台进行沟通，将 STKOS 数据切分成多个小文件，以便系统可多线程并发导入；另一方面也可考虑直接采用纯 XML 数据库对 STKOS 版本数据进行存储与索引。
- (2) STKOS 数据变更信息设计尚待规范化：目前，STKOS 数据加工平台尚无明确关于 STKOS 数据变更信息设计的规范。版本管理系统参照 UMLS 版本数据变更文件格式^[3]，将其简化为：单元素关系变更、二元关系变更两种。单元素关系变更指变更仅仅涉及一个元素，如：STKOS 规范概念或者基础术语的增加或者删除。二元关系变更指涉及到两个元素的变更，是一种关系的变更，如：规范概念与规范概念、基础术语与基础术语之间关系的变更，也有可能是规范概念与基础术语之间关系的变更。考虑到正式 STKOS 发布元数据模型，共定义了 6 类数据对象，即：来源术语、基础术语、STKOS 规范概念、来源词表、范畴表与范畴类^[8]，正式 STKOS 数据变更信息会复杂得多。未来系统版本变更信息设计需要根据正式 STKOS 数据变更信息进行重新优化或设计。
- (3) STKOS 跨版本服务功能有待增强：目前，版本管理系统仅提供跨版本变更信息对照显示功能，只能针对两个版本之间的数据变更信息进行对比显示，尚不能直接反映 STKOS 数据完整的历史变更信息。未来，STKOS 版本数据成熟之后，

版本管理系统需提供更友好的界面展示 STKOS 数据完整的历史变更信息。同时, 需要设计数据接口, 让第三方系统也可以完整地调用 STKOS 数据所有历史变更数据^[9]。

5 结 语

通过设计当前正式服务版本、历史版本、临时版本三种不同的版本类型, 科技知识组织体系版本管理系统可同时对STKOS进行多版本管理和版本内小修订管理, 以便及时向用户提供最新STKOS数据。目前, STKOS版本管理系统已根据医药卫生范畴STKOS数据进行开发与部署, 总体运行状态良好。现有系统框架、数据结构设计等可移植到未来完整STKOS数据的版本管理系统中, 也可为其他类型词表版本管理提供参考借鉴。

参考文献:

- [1] 孙坦,刘峥.面向外文科技文献信息的知识组织体系建设思路[J].图书与情报,2013(1):2-7.
(Sun Tan, Liu Zheng. Methodology Framework of Knowledge Organization System for Scientific & Technological Literature [J]. Library & Information, 2013(1):2-7.)
- [2]王颖. 科技知识组织体系共享服务平台建设. [EB/OL] .(2012) [2014-03-12].
<http://sharestkos.las.ac.cn/>
(Wang Ying. Construction of STKOS Sharing Service Platform. [EB/OL] .(2012) [2014-03-12].
<http://sharestkos.las.ac.cn/>)
- [3] National Library of Medicine. UMLS® Reference Manual. [EB/OL] .(2009)
[2012-10-12].)http://www.ncbi.nlm.nih.gov/books/NBK9685/#ch03.I33_Descriptions_of_Each_File.
- [4] Caracciolo C, Dister S, Johannsen G, et al. VocBench v2.1 User Manual[EB/OL] .[2014-09-02].
http://vocbench.uniroma2.it/documentation/VocBench_v2.1_user_manual.pdf.
- [5] 江梅,田明,孙红为.《CA on CD》光盘数据库检索及应用[J].现代情报,2004(6):147-151.
(Jiang Mei, Tian Ming, Sun Hongwei. Search and Application on CA on CD Database [J]. Modern Information, 2004(6):147-151.)
- [6] Apache Solr. Apache Solr Reference Guide [EB/OL].2014-10-13].
<http://mirror.bit.edu.cn/apache/lucene/solr/ref-guide/apache-solr-ref-guide-4.10.pdf>.
- [7] ISO. Information and Documentation -- Thesauri and Interoperability with Other Vocabularies-Part 1: Thesauri for Information Retrieval[EB/OL] .[2012-12-21].
http://www.iso.org/iso/catalogue_detail.htm?csnumber=53657.
- [8]宋文. STKOS 发布元数据设计[EB/OL].[2014-02-23].
http://168.160.16.186/conference/dome_ch/2012/downloads/pdf%E5%8F%91%E8%A8%80%E7%A8%BF/%E5%AE%8B%E6%96%87_STKOS%E5%8F%91%E5%B8%83%E5%85%83%E6%95%B0%E6%8D%AE%E8%AE%BE%E8%AE%A1.pdf.
(Song Wen. Design of STKOS Publishing Metadata[EB/OL] .[2014-02-23].
http://168.160.16.186/conference/dome_ch/2012/downloads/pdf%E5%8F%91%E8%A8%80%E7%A8%BF/%E5%AE%8B%E6%96%87_STKOS%E5%8F%91%E5%B8%83%E5%85%83%E6%95%B0%E6%8D%AE%E8%AE%BE%E8%AE%A1.pdf.)
- [9] 刘毅,汤怡洁,周子健,等. 科技知识组织体系共享服务平台服务接口建设研究[J].现代图书情报技术, 2014(7-8):9-16.
(Liu Yi, Tang Yijie, Zhou Zijian, et al. Research and Construct of the Service Interface in STKOS Sharing Infrastructure[J].New Technology of Library and Information Service,2014(7-8):9-16.)

(通讯作者：刘丹军，ORCID:0000-0003-1049-0982，E-mail: liudj@clas.ac.cn)

作者贡献声明：

刘丹军，文奕：版本管理系统总体规划、框架流程设计；论文起草。付鸿鹄，胡正银：系统功能设计、版本管理系统数据结构设计。

向彬：STKOS数据版本管理功能实现。

杨宁：版本管理系统技术框架设计及实现；论文修订。钱力：STKOS数据Solr索引功能实现。

刘春江：STKOS版本管理第三方接口功能实现。