

科技知识组织体系协同工作系统实现机制研究*

赵胜钢, 吴思竹, 钱庆, 胡铁军

(中国医学科学院医学信息研究所, 北京, 100020)

摘要: 科技知识组织体系在建设和完善过程中需要跨地区、跨行业专家协同工作, 需要协同工作系统的支持。协同工作系统成功开发依赖于协同工作机制, 包括明确工作流程、标准化功能调用和数据调用的接口。文章探讨了保证协同工作系统正常运行的工作机制、机制实现面临的问题及其基于 OSGi 协议的解决方案。

关键词: 体系结构; 协同工作机制; 知识管理; OSGi

Study on the Collaboration Mechanism of the Science and Technology Knowledge Organization System

ZHAO shenggang, WU sizhu, QIAN qing, HU tiejun

(Institute of Medical Information /Library Beijing 100020)

Abstract: The collaborations between different experts are required during the process of building and optimizing the Science and Technology Knowledge Organization System (STKOS). How to run the collaborations well is a challenge since the experts working on the same system come from different industry domains, and they need work together in different physic locations. This paper analyzes the requirements and the challenges of STKOS, discusses the mechanism of the collaborations and the OSGi based solutions to the questions during the mechanism implementation.

Keywords: Software architecture; Collaboration mechanism; Knowledge management; OSGi

1 引言

信息时代网络技术发展使跨地区、跨行业的专家通过网络进行协同工作, 共同完成一项任务成为可能。协同工作能否顺利进行取决于协同工作机制建立及工作数据共享。数据共享的关键问题是如何在多用户协同操作分布式数据库时保持共享数据的一致性, 化解资源存取冲突, 以及如何解决协同工作系统中多用户协同控制问题。解决问题的途径在于分布在不同子系统的数据库服务器之间以及用户与数据服务器之间建立一个统一的通信接口^[1]。

目前多用户协同工作机制主要使用中间件技术。中间件位于分布式环境之上, 用户应用程序之下。传统的面向对象的技术主要强调封装与继承性, 对象一般都只存在一个程序中, 而程序外界都无法访问该对象, 在分布式环境中分布式对象不仅具有传统面向对象封装的特征, 而且更重要的是分布式对象的分布性可以分布在网络中任何一个地方^[1]。

科技知识组织体系 (Science and Technology Knowledge Organization System, STKOS), 是面向科技领域的, 对科技领域知识架构进行表达和有组织的阐述的一种语义工具, 由英文超级科技词表和本体两部分组成。建立科技知识组织体系是国家“十二五”科技支撑计划的研究课题之一。

STKOS 的建立和完善过程需要跨地区、跨行业的专家协同工作, 是一个典型的分布式多用户协同工作过程。STKOS 协同工作系统是支持这种协同工作的软件工具, 是支持

*本项目为国家科技支撑计划“科技知识组织体系的协同工作系统和辅助工具开发”(编号: 2011BAH10B02)研究成果

STKOS 建立的集素材、超级科技词表和本体构建与管理为一体的多层次的协同工作系统,除了提供用户管理、任务分配、用户日志管理、并发控制、统计管理、配置管理等公用功能外,还面向不同层面内容建设提供不同支持服务功能。

STKOS 协作工作系统主要用户是在位于不同地域,从事不同领域研究的专家,因此在系统设计时,面临着协同工作实现方式的挑战,即如何使系统以具有用户群体性、交互性、分布性的协同工作方式,便于地域分散的群体能够借助计算机网络技术,共同协作来完成一项任务,从而减小或消除人们在时间和空间上的障碍,提高群体工作的质量和效率。由于 STKOS 系统建设是一个渐进过程,因此,系统建设首先需要有一个合适的工作流程以及和工作流需求相适宜的技术解决方案^[2]。解决问题的思路是在传统分布式系统基础上,利用成熟的中间件技术与标准数据交换协议,面向 STKOS 内容建设需要,完成网络化的 STKOS 工作流设计,并研究适用于 STKOS 建设与管理、支持网络化工作流的分布与集中相结合的计算机支持的协同工作机制。

本文在深入分析 STKOS 建设需求的基础上,提出网络化的 STKOS 工作流设计方案,提出计算机支持的 STKOS 系统协同工作机制以及基于 OSGi 协议的机制实现技术解决方案。

2 STKOS 协同工作系统需求和基本功能

STKOS 的基本组成元素是科技词汇及其相互之间的关系,STKOS 的建设和完善过程是科技词汇的收集、遴选和对科技词汇之间的关系进行甄别梳理的过程。这些科技词汇覆盖了各个行业的专业领域,对这些词汇进行遴选、对之间关系进行判定是一个知识性很强的任务,需要来自不同领域相关行业的多位专家的协同工作。这种协同工作必须要有一个在线系统进行支持,这个系统需要能满足这些不同领域的专家共同对同一个或一组词汇进行协作加工的要求,还要面临因这些专家的工作时间和地点不一致带来的对协同工作的挑战。

建立 STKOS 协同工作系统的目的是提高 STKOS 的建设效率和质量。STKOS 协同工作系统是一个支持 STKOS 素材、超级科技词表、领域本体和科研本体内容建设的协同工作环境和应用工具集。其中,协同工作环境是一个数据在线加工与管理系统,即用户可以通过浏览器登录系统,进行数据加工,功能除了提供用户管理、任务分配、用户日志管理、并发控制、统计管理、配置管理等公用功能,还面向不同层面内容建设提供不同支持服务功能:1) 面向 STKOS 素材建设提供多来源异构素材的格式转换、规范化批量导入与导出、检索与浏览、统计分析、编辑修改、自定义表架构、关系/属性类型新词推荐等功能;2) 面向超级科技词表内容建设将提供 3 个层面的服务,即基础词库层、规范概念层与范畴体系层,将具有对术语和各类概念与范畴进行评估、筛选、映射、关联、加工、集成和管理的能力,支持概念、概念关系、概念属性、范畴(概念语义)、概念—词语映射、范畴—概念映射等知识单元的协同构建,支持一致性检验,并可按标准的本体描述语言将规范词表系统导出供其他系统使用;3) 面向本体(包括科研本体和领域本体)建设提供本体的创建、编辑、浏览和维护功能,支持本体的映射和集成,支持通用的本体查询语言进行本体的查询和推理,支持本体演化管理等;应用工具集为一些可离线使用的客户端小工具,允许用户在没有互联网连接的情况下,使用离线数据加工工具,处理加工一些数据,然后转化为系统可以接受的标准数据交换格式,导入到系统。

实现上述系统功能的难点是协同工作机制和计算机支持的协同构建技术,即对于相互关联程度比较高的知识单元,如何在不影响用户体验的条件下,保证用户无障碍地协同构建知识。对协同工作机制进行研究分析的前提是对 STKOS 的协同工作流程进行分析研究和定义。

3 STKOS 协同工作流分析

STKOS 协同工作系统是一个支持不同角色的用户在各个层次共同进行数据加工的过程

程,从技术角度看是支持 STKOS 协同工作流程的定义和运行。工作流程涉及两个关键问题,一是管理问题,解决谁该干什么的问题,解决思路是通过建立公用的工作流程和管理引擎,随时定义和调整 workflow。另一个是技术问题,即如何解决多人协同工作时的资源冲突问题,尤其是对多人同时对同一数据发出编辑(含新建和删除)请求时可能产生的冲突。STKOS 协同工作系统的工作流程设计必须能同时解决好上述两个关键问题。本文提出具体的解决方案:

(1)、分层循环工作流程

图 1 描述了 workflow 设计方案,即以数据共享为核心的分层循环工作流程。

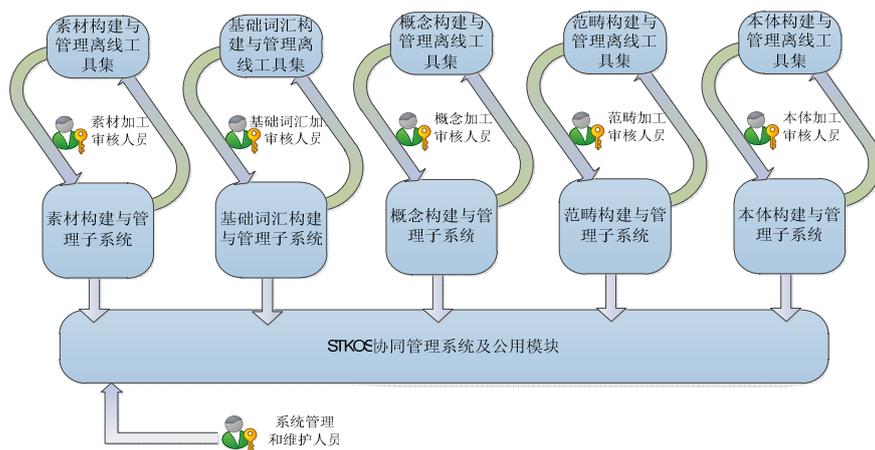


图 1 分层循环 workflow

在此 workflow 中,协同工作系统的管理系统提供 workflow 定义和维护机制以及系统插件的管理。各个层次的子系统以插件方式集成到管理系统中,获取各自系统的用户与权限定义以及 workflow 的定义,并以此为基础开展各自层次的数据循环加工工作,各子系统之间的交互通过管理系统的接口管理功能实现。各子系统内部的工作流程由各子系统分别定义实现。

这个 workflow 的设计,一方面满足了不同子系统之间的集成、交互和数据交换的要求,另一方面也满足了在对 workflow 总体定义、对用户权限总体管理的同时,各个子系统的工作流程能够闭环运行,满足彼此低耦合的要求。

(2)、 workflow 资源冲突解决机制

解决 workflow 资源冲突问题主要是设计实现一个基于数据记录锁定的协同数据处理流程,如图 2 所示。

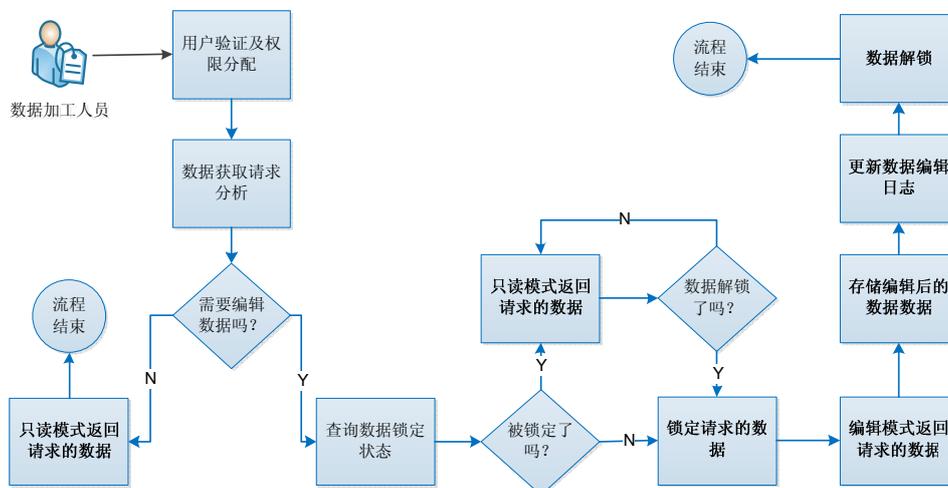


图 2 workflow 冲突解决机制

在图 2 所示流程中, 数据加工人员(包含素材、基础词汇、规范概念、本体和范畴各级加工人员)通过用户验证获取了相应操作权限后, 对协同工作系统发出了数据获取请求, 系统对此请求进行分析, 判断是否包含数据编辑要求, 如果不包含(只是查询, 此时不会产生资源访问冲突), 则以只读的模式将用户需要的数据返回给用户; 如果判定包含数据编辑要求, 则先查询该数据的锁定状态, 看是否被其他用户锁定, 如果没有被其他用户锁定, 则立即锁定该数据, 如果已经被其他用户锁定, 则以只读的方式返回数据, 同时通知用户等待数据解锁, 经过限定时间的等待后, 如果其它用户解锁, 则立即锁定该数据; 数据被成功锁定后, 系统以可编辑模式返回用户需求数据, 待用户提交所进行的操作之后, 存储所编辑的数据, 同时更新数据变化日志, 数据解锁, 完成工作流程。

4 STKOS 协同工作机制设计

STKOS 协同工作机制设计要求是: 在传统分布式系统基础上, 面向 STKOS 内容建设需要, 支持网络化的 STKOS 协同 workflow, 研究适用于 STKOS 建设与管理的、分布与集中相结合的计算机支持的协同模式, 满足群体性、交互性、分布性协同工作方式的客观需求, 便于地域分散的群体借助计算机网络技术, 共同协作来完成一项任务, 减小或消除人们在时间和空间上的障碍, 提高群体工作的质量和效率。

通过对协同工作系统的需求分析和 workflow 分析可以看出, 协同工作系统的工作过程从功能上看是不同角色的系统用户针对处于不同状态的“词表”数据进行加工的过程, 从信息流上看是“词表”数据在各个状态之间的变化过程及对其属性(包括关系属性)的归并、定义的过程。“词表”状态的变化有其内在的规律性, 加工过程有对其步骤设计的科学性合理性要求, 同时加工过程有群体性、交互性、分布性协同工作方式的客观需求, 这三点决定了对协同工作系统工作机制进行研究的必要性。

STKOS 协同工作系统分为离线工具集和在线系统两个部分, 涉及到协同工作的部分主要是在线部分。由于 STKOS 协同系统的各个子系统由不同的参与单位分别开发, 在开发阶段存在着各个应用子系统的异地分布式开发的问题; 素材数据库、基础词库、规范概念词库、范畴体系库和本体数据库在物理上分布在不同的地点, 同时数据架构设计也可能存在差异; 其次, 在整个开发与部署阶段, 这些数据库之间存在着共享与支持关系, 在应用程序运行过程中模块之间存在共享(复用)和调用关系; 第三, 在应用阶段, 各个子系统存在着统一部署、功能集成等要求, 包括系统的用户和权限需要统一管理, 加工流程需要统一定义与维护, 各个子系统也需要共享用户管理数据和流程管理数据等。因此协同工作机制要解决的主要问题是: 如何解决分布式系统和异地异构的数据库之间的数据存取和数据共享问题, 和由不同开发部门开发出的系统功能模块的复用和集成问题。解决这一问题的关键是数据存取接口标准化和功能集成接口标准化的问题。针对不同阶段的问题, 模式设计分为开发阶段工作模式和应用阶段工作模式。

(1) 开发阶段工作模式

图 3 描述了开发阶段工作模式。在此模式下, 各个开发单位各自部署和管理自己的数据库, 在各自的开发环境中进行开发工作, 各个单位须按照统一的功能调用规范(本设计采用了 OSGi 规范)提供对可复用功能和各自数据库操作功能的接口。各个单位通过调用通用数据接口(本设计采用了 OSGi 规范的数据接口)的方式访问其他子系统的数据库, 共享可复用的功能模块。

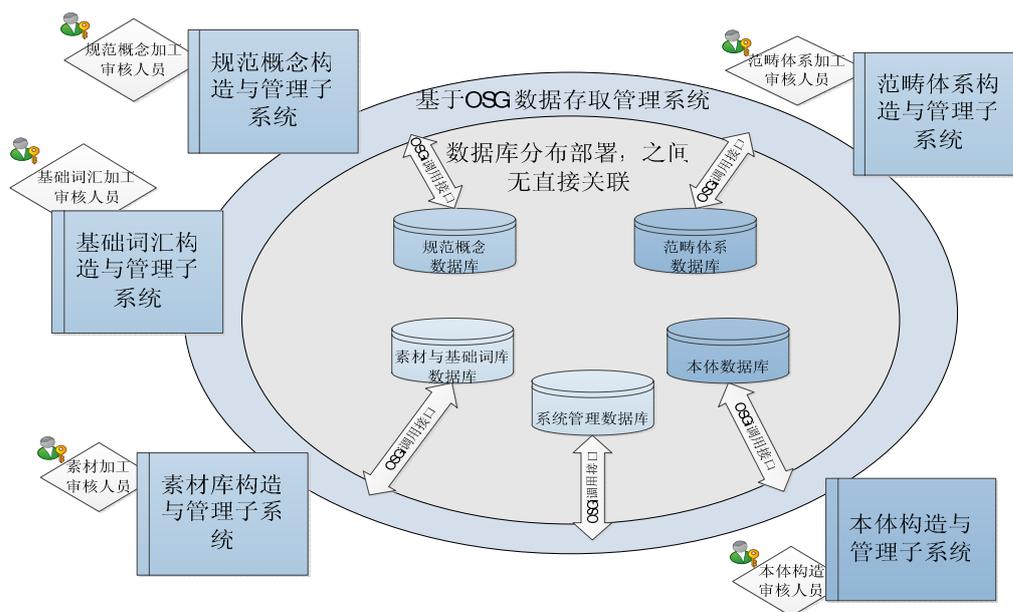


图 3 协同系统开发工作模式

(2) 应用阶段工作模式

图 4 描述了应用阶段工作模式，在此模式下，各个数据库和各个子系统的应用程序统一部署在中心服务器中，其数据之间以及功能之间的共享与调用仍然通过OSGi接口模式进行，除了系统配置上的不同外，数据库结构与子系统功能设计实现与开发阶段并无不同。

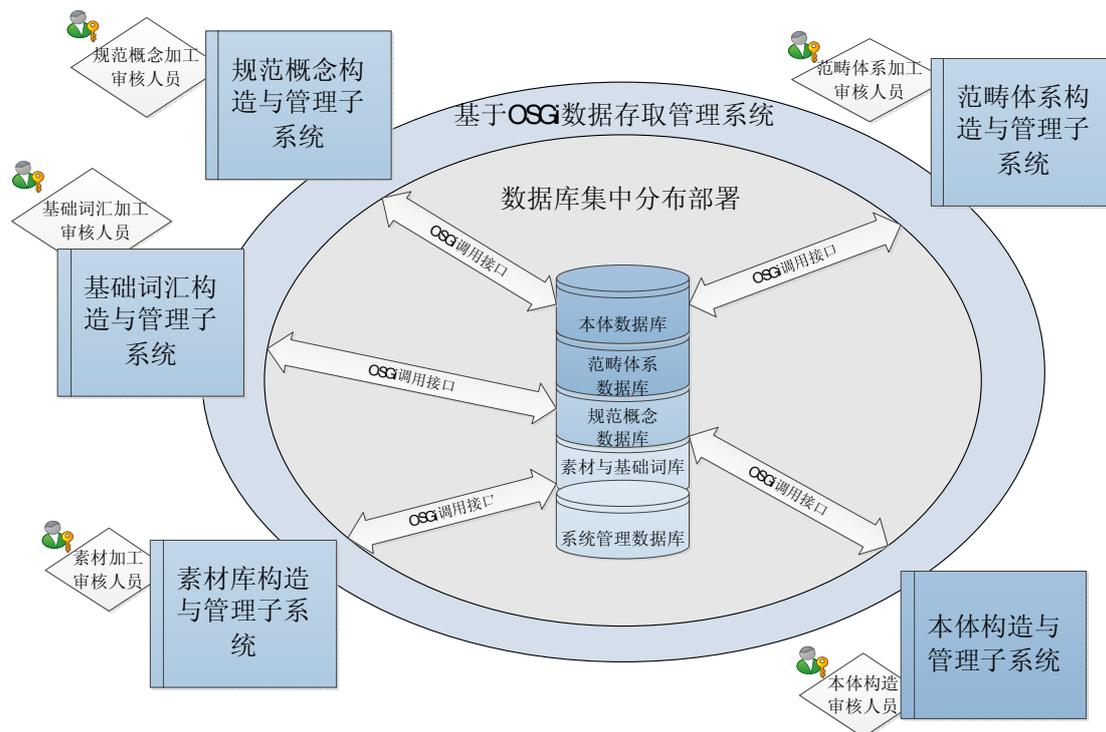


图 4 协同工作系统应用工作模式

5 技术解决方案

STKOS 协同工作系统加工对象包括素材、超级科技词表和本体等各种知识架构，每种

知识架构的组织方式、加工方式和表现方式各有特点；同时，STKOS 协同工作系统具有“分布开发、集中部署”的要求，即开发阶段各个开发单位的开发地点、开发环境和开发方法相对独立，但各个开发单位开发的子系统之间的功能上和数据上有着内在的联系，最后开发完成的各个子系统需要无缝集成在一起，集中安装部署到统一的服务器上。

STKOS 协同工作系统分为 6 大子系统，各个子系统之间的交互关系可以分为数据共享关系和功能交互关系两大类。交互关系的实现取决于数据共享和功能交换接口的标准化。

(1) 数据共享接口标准化解决方案

STKOS 协同工作系统在开发阶段和应用阶段都涉及到数据共享和数据调用，数据共享关系主要包括：上层系统调用下层数据，综合系统调用各层数据，如用户和权限表的共用和概念层对基础词汇层数据调用。

开发阶段的分布式特性导致了数据共享的复杂性，各个开发点需及时获取其他开发点的数据结构和数据，同时每个开发点的数据结构和数据的变化不应影响到其他开发点的开发，提供数据共享的开发点开发过程也不应该受到数据共享的影响。

可能的数据共享解决方案包括数据库复制模式、远程直接调用模式、中心数据库模式、分布数据库+WebService 模式和分布数据库+OSGI 封装模式。

1) 数据库复制模式：每个开发点都复制其他开发点的数据结构和数据到本地数据库。

优点：开发效率高

缺点：数据结构和数据的变化对开发影响较大、数据库同步管理比较繁琐

2) 远程直接调用模式：各个开发点提供数据库远程调用接口。

优点：免除数据库同步。

缺点：数据结构和数据的变化对开发影响较大、数据传输的效率影响开发效率、数据安全性较低。

3) 中心数据库模式：设立中心数据库服务器，各个开发以中心数据库为中心进行开发。

优点：免除数据库同步、安全性较好。

缺点：数据架构和数据的变化对开发影响较大、数据传输的效率影响开发效率

4) 分布数据库+WebService 模式。

优点：免除数据库同步、安全性较好、数据架构和数据的变化对开发无影响，便于系统开发过程中和系统运行过程中的维护。

缺点：WebService 接口的设计质量与数据传输的效率影响开发效率。

5) 分布数据库+OSGi 封装模式。

优点：免除数据库同步、安全性较好、数据架构和数据的变化对开发无影响，有很强的灵活性和扩展性，支持软构件热插拔的概念和实现。

缺点：接口的设计质量与数据传输的效率影响开发效率，开发环境局限于 java 平台，开发门槛较高。

通过对各种数据共享模式的特点分析，建议采用分布数据库+OSGI 封装模式。

(2) 功能集成接口标准化解决方案

功能交互关系主要包括功能的集成和相互调用，在功能集成方面，主要包括用户与权限管理功能、工作流程动态分配与管理功能、系统管理与数据备份功能、综合查询与应用等功能的集成，同时还要考虑到灵活性与扩展性要求，满足“通用的、开放的、可扩展的、能够提供统一标准接口的、分层次”的要求和“自适应、可进化、模块化设计、稳定、健壮的系统内核、支持各模块的动态热拔插”等特征。

STKOS 协同工作系统在开发阶段涉及到公用模块调用，部署阶段和应用阶段涉及到功能集成。分布式开发对模块共享和版本控制形成挑战，系统集中部署对模块集成接口提出要求。

可能的模块共享和集成解决方案包括采用 DLL、OAI、OSGi 等接口协议作为定制开发的标准。通过对以上各种模块共享和集成模式的特点分析,建议采用 OSGi 协议、规范模块接口。

优点:提供热插拔式系统集成,模块间耦合性小,兼容性强。

缺点:开发难度较大,开发语言局限于 JAVA。

(3) OSGi 规范及应用

如前所述,通过对数据共享和模块集成的各种技术分析,OSGi 规范被采用作为数据共享和功能集成的技术解决方案。

OSGi (Open Service Gateway Initiative) 是建立可互操作的分布式应用程序的规范,它定义了应用程序如何在 Web 上实现互操作性。OSGi 规范由 OSGi 联盟于 1999 年制定,也称为 Java 语言动态模块系统。该规范使得组件间可以相互隐藏实现的细节,通过服务来进行通信。它提高了组件的重用率,降低了开发的复杂度,使组件易于调配和管理,提供了一个通用软件运行的平台。

STKOS 协同工作系统和辅助工具的开发过程中,应用 OSGi 规范作为功能集成接口标准有助于系统中各个子系统及辅助工具的功能模块的合理划分,使系统具有高内聚、低耦合的特征,组件更加灵活易用;同时支持模块化的系统开发模式,便于 STKOS 协同工作系统各功能模块及工具的开发、调用、组配、部署和管理;应用 OSGi 规范作为数据共享接口标准使数据调用接口透明化,消除了因数据库不同地域分布可能引起的数据交流障碍;在系统部署和动态升级方面,OSGi 支持 bundle 动态部署,无需停机重启,就可以实现系统内部模块功能的动态更新,只要保持接口不变,就可以实现对系统及外部功能交换无影响,这样即便 STKOS 协同工作系统内部模块进行更新或升级也不会影响到用户的系统使用。在具体实现上,基于 OSGi 规范,系统组件内部的模块通过一个具有动态加载类功能的微内核进行连接,将不同模块(bundle)之间的类接口采用包引用或服务注册的方式进行管理,实现模块间的隐私保护和公共功能的调用及共享(比如数据访问模块、日志管理模块等),确保了模块之间的相互协作能力,也有助于模块的复用,提高了利用率。此外,OSGi 规范与成熟的基础框架如 Equinox、Filex、Spring DM 等的结合使用,有助于对模块的存储、管理及配置。因此在 STKOS 协同工作系统和辅助工具的开发中使用 OSGi 规范能够满足项目系统构建的“自适应、可进化、模块化设计、稳定、健壮的系统内核、支持各模块的动态热拔插”的开发与应用要求。

6 结束语

STKOS 是一个支持对海量词汇数据进行加工、处理的知识组织体系,该体系的构建是一个涉及多方面因素密切协同的系统工程。作为实现这一工程的支持系统,STKOS 协同工作系统的协同工作模式设计所应用的方法和实现技术均面临着新的挑战。本文提出的协同工作系统的工作流设计方案、协同工作机制以及基于 OSGi 的数据共享和功能集成的解决方案,是对如何解决协同工作系统面临问题的一种尝试,其中,将 OSGi 规范应用到协同工作系统的设计中,是解决此类问题的一个创新。

目前 STOKS 协同工作系统正在建设过程中,本文描述的设计与解决方案是否可行,还有待于在实践中得到验证,这些设计与方案必然会随着项目的推进而面临的需求变化而不断优化和完善。要真正实现设计方案还需要大量细致的工作要做,如支持 OSGi 协议的数据库设计和实施规范,如何定义、实现基于 OSGi 的模块接口,基于 OSGi 接口的数据存取冲突机制的详细方案等,都需要进行进一步研究。

参考文献

- [1] 郭朝珍, 赖诗长, 庄苗. 政务信息共享平台中多用户协同工作机制的研究[J]. 系统仿真学报 Vol. 15 No. 1 Jan. 2003 72-75
- [2] 吴汝明, 辛小霞. 基于协同工作的高校校务管理系统[J]. 中山大学学报 2009 年 3 月第 48 卷增刊 113-116

作者介绍

赵胜钢, 1967 年生, 研究方向为知识管理与软件系统工程

通信地址: 中国医学科学院医学信息研究所

联系电话: 82128768 邮箱: zhao.shenggang@imicams.ac.cn

科技知识组织体系协同工作系统实现机制研究

作者: [赵胜钢](#), [吴思竹](#), [钱庆](#), [胡铁军](#)
作者单位: [中国医学科学院医学信息研究所, 北京, 100020](#)

引用本文格式: [赵胜钢](#). [吴思竹](#). [钱庆](#). [胡铁军](#). [科技知识组织体系协同工作系统实现机制研究](#) [会议论文] 2012