

水稻领域知识服务系统的设计与实现*

□ 寇远涛 赵瑞雪 鲜国建 朱亮 / 中国农业科学院农业信息研究所 北京 100081

摘要: 随着科研方式的转变, 科研用户对专业化、知识化服务的需求不断增强, 因此面向专业领域构建集文献检索获取、领域知识导航、知识管理与共享和科研交流与合作等功能服务为一体的领域知识服务系统, 成为信息服务机构的必然选择。文章提出了水稻领域知识服务系统的功能模型, 并开发实现了系统主要功能。

关键词: 水稻, 领域知识服务, 知识组织, 知识地图

DOI: 10.3772/j.issn.1673—2286.2012.12.008

1 引言

随着科研环境与科研方式的变化, 科研用户对专业化、知识化服务的需求不断增强, 因此面向专业领域构建集文献检索获取、领域知识导航、知识管理与共享和科研交流与合作等功能服务为一体的领域知识服务系统, 成为信息服务机构的必然选择。国内外在这方面的研究和探索已经出现了不少成果, 如美国的研究者合作网络(VIVO)^[1-3]、英国的虚拟研究环境(VRE)^[4,5]、中国科学院国家科学图书馆的专业领域知识环境(SKE)^[6]、中国农业科学院国家农业图书馆面向学科的数字图书馆^[7], 为本文研究的开展提供了借鉴。

水稻是我国最主要的粮食作物, 其种植面积居我国粮食作物的首位, 全国有65%以上的人口以大米为主食。稻米产量在很大程度上制约着我国的粮食市场走势和粮食安全形势^[8]。因此我国非常重视水稻相关研究。为了更好地为水稻领域科研工作者提供信息服务, 本文提出了领域知识服务系统的功能模型, 并面向水稻领域设计开发了知识检索获取、知识关联、知识地图和知识组织与管理等系统功能, 实现了水稻领域知识服务系统。

2 系统功能模型

水稻领域知识服务系统通过对各类科技信息资源的分析与挖掘, 从中发现水稻领域知识, 并将这些

知识经识别、理解、筛选、格式化抽取形成水稻领域知识库。在知识库基础上, 对水稻领域内的主题、机构、人员、文献等知识对象进行关联、集成, 实现科技资源的知识检索获取、知识关联推荐、知识地图、知识组织与管理等功能服务。

水稻领域知识服务系统的功能模型如图1所示, 包括:

(1) 知识应用功能, 主要提供领域知识服务系统的核心服务, 包括知识导航、知识检索获取、知识关联、知识地图、知识共享、知识管理等功能。

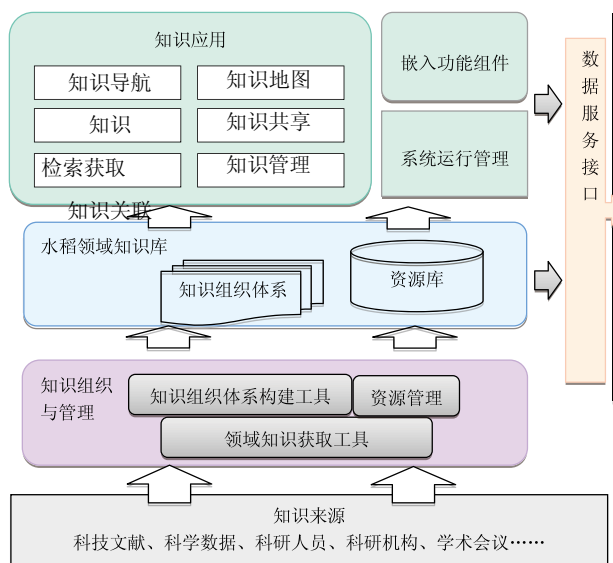


图1 水稻领域知识服务系统功能模型

* 本文系国家“十二五”科技支撑计划项目课题“基于STKOS的知识服务应用示范”(编号: 2011BAH10B06)的研究成果之一。

(2) 知识组织与管理功能,包括领域知识获取工具、知识组织体系构建工具和资源管理功能,实现从各种类型的知识来源到资源库的加工与转化,科研本体和领域本体类、属性等的管理维护,以及各种资源的管理,完成水稻领域知识库的建设。

(3) 嵌入功能组件,主要提供科研交流与协作、机构知识库、科技监测等功能支持,功能相对独立,采用组件式嵌入或链接接入,为科研人员提供在线交流与知识共享环境。

(4) 数据服务接口,主要实现领域知识服务系统的可扩展、可延伸性,支持对系统资源以及服务功能的第三方系统调用,扩展系统服务方式。

(5) 系统运行管理,系统管理、用户管理等支持系统正常运行的基本功能。

3 系统设计与实现

3.1 技术选型

本系统采用Java为主要开发语言,基于J2EE架构,平台环境为“Windows2008+Tomcat6.0+Mysql5.5”,开发工具为MyEclipse。针对本系统大容量数据的特点,采用分布式结构,有利于分担系统计算和存储压力。

在水稻知识库构建中,将本体的理念和技术引入资源组织中,知识库采用O-R(本体-关系)映射技术实现对资源的动态存储,采用Hessian技术实现知识库跨语言、跨平台的调用,更具备开放性。

系统前台知识应用采用SSH架构,能够从知识库搭建的知识结构动态调整栏目及信息项目,以实现跨领域的应用,并运用AJAX、RSS、GIS技术、知识可视化技术等相关软件技术实现知识服务功能。系统后台知识组织与管理功能采用“仿桌面”技术,实现了Web系统下的Windows操作体验。

3.2 核心功能实现

以下给出知识检索获取、知识关联、知识地图和知识组织与管理等系统核心功能的设计与实现。

3.2.1 知识检索获取

在知识检索上,基于本体应用,提供了如下两种

智能检索扩展:①上下位、同义词、相关词扩展,根据主题概念之间上下位、同义词、相关词的关系,可以实现检索词的扩展,从而获得更全面的检索结果。同时点击扩展到的主题词还可以把它作为新的检索词进行检索。②中英文双语扩展,通过概念的中英文映射,实现检索词的扩展,从而实现双语检索。

知识的全文获取是进行知识检索的终极目标。系统建设时利用已有的知识库资源,进行知识的二次利用和关联发现,实现了基于知识库的情景敏感,配合知识检索实现目标资源全文的获取,最大程度满足用户需求。图2给出了用户情景敏感本体的原理示意图。情景敏感本质上是描述用户信息,解决用户在什么情况下使用和接受服务,可以获取什么样的服务和资源这一问题。结合扩展的知识库系统,可以根据获取的用户信息,提供更多的扩展信息,达到一站式全方位服务的目的。

3.2.2 知识关联

领域知识关联是将各种信息载体中的共现信息定量化的分析方法,以揭示信息的内容关联和特征项所隐含的寓意。主要包括基于内容相似度的知识关联、基于研究人员的知识关联、基于研究机构的知识关联和基于研究主题的知识关联。

(1) 基于内容相似度的知识关联

对资源的某些内容特征进行分析,形成资源的文本特征索引,然后对不同资源间的索引特征向量求相似度,形成相似资源矩阵数据库。当用户浏览某篇文献时,系统即可从此相似资源矩阵中获取其他关联资源,推荐给用户。这就需要根据文献的摘要、主题词和正文信息进行建模。而这些信息最终都是由农业词表或农业本体中的概念来构成,这样每个文献模型就是一个向量,该向量的每个分量就是主题词表中的词,这样我们就可以通过计算两个向量的夹角的余弦来判断两个文献的相似度^[9]。

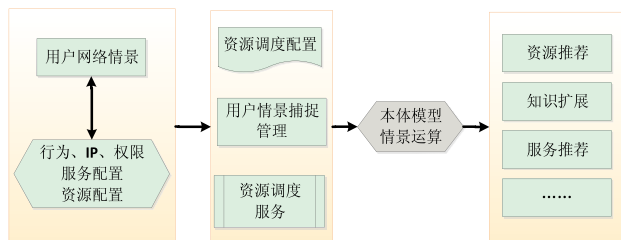


图2 情景敏感示意图

(2) 基于研究人员的知识关联

向用户推荐与其当前浏览文献的作者相关的其他作者,采用以下三种方式来实现:①基于文献作者的合作、指导等关系,形成作者之间的关联并基于此向用户推荐与当前浏览文献的作者有合作或者指导关系的其他作者;②基于共词分析的作者关联推荐,若两个作者发表的论文中都含有同一关键词则认为是关联作者;③基于引证分析的作者关联推荐,依据是相互引证的多篇文献作者之间存在的关联关系。

(3) 基于研究机构的知识关联

向用户推荐与其当前浏览文献的研究机构相关的其他机构。从研究机构之间的合著关系、共词分析和引证分析三个方面来形成研究机构之间的关联,基于此关系实现研究机构关联推荐。机构知识关联的实现与研究人员的关联实现基本方法一致。

(4) 基于研究主题的知识关联

向用户推荐与其当前使用的检索关键词所属研究主题相关联的其他主题关键词。依据领域知识网络(领域本体),推荐出邻近概念的相关词汇,从两个方面推荐:领域概念和科技概念。领域概念包括与概念相关的专家、研究机构、法规等;科技概念包括用、代、属、分、参的邻近科技词汇。

3.2.3 知识地图

知识地图以图形化的方式直观地展示专业领域知识的组成结构、知识关联、发展变化等信息,帮助用户快速、准确地对所需要的知识进行定位和获取。以下从地图的生成与展示、地图节点生成、地图节点展示算法、知识定位和获取四个方面介绍本文的实现思路。

(1) 地图生成与展示

本文在处理图结构时,借用物理学的原理,模拟带电荷的质点表示节点,用弹簧表示弧,节点之间带有同种电荷相斥而远离,用弹簧约束各个节点,如此整个图结构内的元素就可以在有限的空间里合理自然地分布和展示。地图的生成与展示以Flash技术为展现形式,直观动态地生成和展示知识地图。

(2) 地图节点生成

节点具有质量,在力的作用下会在空间中产生加速度而移动。同时,所有的节点都具有同种电荷,会相互排斥,避免相互重叠。在应用中,每个节点可以表现为一个可交互操作的面板,可在其上展示该节点的所有内容,

例如标题、文字描述,甚至图片、视频和音频。节点的外观大小和质量可以用来表示节点的权值。

(3) 地图节点展示算法

如图4所示的基本情形,节点A和节点B通过弹簧相连。在某时刻,设A和B的坐标分别为 r_A 和 r_B ,都为二维向量;A和B的质量分别为 m_A 和 m_B ,电量分别为 q_A 和 q_B 。这个位置并非一定是平衡位置,且它们正在以各自的速度运动中。

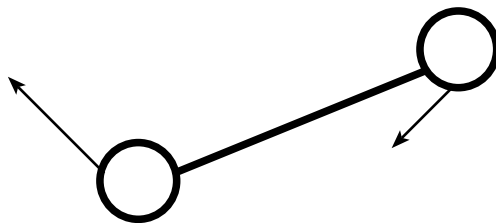


图4 地图节点展示算法

根据静电斥力的向量形式计算公式,A受到来自B的静电斥力为:

$$F_A = k_e \frac{(q_A q_B (r_A - r_B))}{|r_A - r_B|^3}$$

其中 k_e 为静电力常量,可以设置为真实世界中的 $8.987 \times 10^9 \text{ N} \cdot \text{m}^2 \cdot \text{C}^{-2}$,也可以根据需要设置为其他数值。相应地,B受到来自A的静电斥力与之相反:

$$F_B = -F_A$$

根据胡克定律,弹簧此时的张力的大小为:

$$T = -k (|r_A - r_B| - L_0)$$

其中, L_0 为弹簧的原始长度。因此,A和B应当受到的弹簧作用力为:

$$T_A = T \frac{r_A - r_B}{|r_A - r_B|}$$

$$T_B = -T_A$$

如此,已经得到了A和B所有所受的力,因而可以求得A和B各自的合力,进而可以求得它们各自的加速度:

$$a_A = \frac{F_A + T_A}{m_A}, \quad a_B = \frac{F_B + T_B}{m_B}$$

则A和B在下一个时刻的运动速度可利用显式积分法求得:

$$v'_A = v_A + a_A \quad t, \quad v'_B = v_B + a_B \quad t$$

A和B在下一个时刻的空间位置也可积分求得:

$$r'_A = r_A + v'_A \quad t, \quad r'_B = r_B + v'_B \quad t$$

(4) 知识定位和获取

通过上述的算法,节点在形成知识地图后都带有相应的标识和位置信息,Flash支持以多种方式和后台服务端进行通信,因此在用户点击某个知识节点后,通过触发通信交互从服务端获得该知识点的详细信息,可用于直接展示在页面上,也可以通过知识地图进入下一步交互程序。

3.2.4 知识组织与管理

知识组织与管理工具实现领域知识内容管理。知识内容可以分为由科研本体和领域本体框架结构和实例库构成的知识组织体系和由科技信息资源构成的资源库。知识组织体系维护工具实现本体框架结构的维护以及本体实例的填充,资源库管理工具实现资源的管理,领域知识获取工具完成从知识来源到资源库的知识抽取与转换。

(1) 知识组织体系维护工具

知识组织体系维护工具主要包括两方面的功能:

①知识组织体系框架管理:能够对知识组织体系本体框架进行管理和维护,具体包括类管理、对象属性管理和数据属性管理。

②知识内容管理:系统提供构建工具,对现有的数据库资源、互联网资源以及文本资源进行抽取和加工。在加工的过程中,按照规范的知识编辑加工工作流程进行质量控制,每条知识都带有机或加工人员的标识,便于对加工内容进行回溯管理和统计。同时对已加工的知识内容,提供方便快速的预览模式,并可以图形化的方式展示编辑后的关联关系,同时对整个加工内容提供统计和管理。

(2) 资源库管理工具

资源库管理工具实现资源的管理,包括编辑维护、导入导出、资源统计等功能。

(3) 领域知识获取工具

领域知识获取工具支持从各种常用关系型数据库中水稻领域知识的自动获取,支持从各种馆藏资源库中进行自动知识获取,并建立知识关联。采用C/S架构,服务器端负责定制所需要采集的数据并通过网络下发到客户端;客户端只需要根据此采集清单,在图形化配置界面的引导下,为采集清单和本地信息系统的数

据库做一个简单的映射后,便可以开始数据采集的整个流程,实时、准确地将数据以统一的格式上报至服务器端汇总处理。

3.3 系统部署

在以上分析设计的基础上,部署系统功能,构建科研本体和水稻领域本体,进行数据的搜集、装载、测试,最终实现了水稻领域知识服务系统。

图5为系统首页,包括领域新闻动态、科研项目、学术活动、资源检索、资源导航、专家信息、科研机构、专业期刊、政策法规等栏目,并对领域内热门主题以标签云的形式可视化展示。

图6为一站式智能检索页面,实现对各类资源的统一检索,可以直接智能命中科研人物。

图7为文献详细信息页,实现了知识关联、知识全文获取等功能,对文献的作者、母体文献做了知识链接,并提供相关资源、共词作者、原文下载等功能。

4 结语

本文提出了水稻领域知识服务系统的功能模型,



图5 水稻知识服务系统首页



图6 智能检索页面



图7 文献详细信息页

从技术层面探讨了系统的开发实现,但还有许多问题需要在后续工作中考虑:(1)需要与水稻专家研究和进一步明确领域本体构建的范围与粒度;(2)资源持续更新的相关机制的建立;(3)领域知识服务的方式的扩展,比如关联数据发布。总之,面向专业领域构建知识服务平台,是提高学科化、知识化服务能力和水平的有效手段,信息服务机构应予以重视并积极推进,为用户的科研创新提供知识支撑。

参考文献

- [1] KRAFFT D B, CAPPADONA N A, CARUSO B, et al. VIVO: Enabling National Networking of Scientists [C]// Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.
- [2] DEVARE M, CORSON J, CARUSO B, et al. VIVO Connecting People, Creating a Virtual Life Sciences Community [J]. D-Lib Magazine, 2007,13(7-8).
- [3] Page 1 of 11 comparison with VIVO [EB/OL]. [2012-01-23]. http://sw1.slis.indiana.edu/repository/systemdoc_files/comparison_with_VIVO%2012-01.pdf.
- [4] ALLAN R, ALLDEN A, BOYD D, et al. Roadmap for a UK Virtual Research Environment [R]. Report of the JCSR VRE Working Group, 2004.
- [5] YANG X, ALLAN R. Web-Based Virtual Research Environments (VRE): Support Collaboration in e-Science [C]// Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology. Washington, DC, USA: IEEE Computer Society, 2006: 184-187.
- [6] 宋文,张士男.专业领域知识环境建设的理念与实践[C]//数字图书馆高层论坛2010年年会论文集.北京:国家科技图书文献中心,2010:19-23.
- [7] 赵瑞雪.数字环境下面向专业研究所的知识服务探索[J].情报杂志,2011,30(6):189-193.
- [8] 傅强,黄世文.水稻病虫害诊断与防治原色图谱[M].北京:金盾出版社,2011.
- [9] 方安,李亚子.虚拟医疗社区中用户相似度计算方法[J].医学信息学杂志,2011,32(3):48-51.

作者简介

寇远涛 (1982-), 博士, 助理研究员, 研究方向: 数字图书馆理论与技术、信息管理与信息系统。E-mail: kyt@mail.caas.net.cn

赵瑞雪 (1968-), 博士, 研究员, 博士生导师, 研究方向: 信息管理与信息系统、信息资源管理、知识组织及数字图书馆。E-mail: zhaorx@mail.caas.net.cn

Design and Implementation of Rice Domain Knowledge Service System

Kou Yuan Tao, Zhao Ruixue, Xian Guojian, Zhu Liang / Agricultural Information Institute of CAAS, Beijing, 100081

Abstract: As scientific research way changes, scientific users' demand of specialized, knowledge-based service grows constantly. So construction of domain knowledge service system with the integration of such services as literature retrieval and acquisition, domain knowledge navigation, knowledge management and sharing, as well as scientific research exchange and cooperation, becomes the inevitable choice of information service. This paper put forward the function model of rice domain knowledge service system, and realized the main system function.

Keywords: Rice, Domain knowledge services, Knowledge organization, Knowledge map

(收稿日期: 2012-08-01)