

大数据时代开放式农业信息知识库构建研究

陈娉婷¹, 官波¹, 沈祥成¹, 吴亚玲¹, 马海荣¹, 邓丹丹², 张月婷², 徐桂梅², 罗治情^{1*}

(1. 湖北省农业科学院农业经济技术研究所, 武汉 430063; 2. 武汉光谷信息技术股份有限公司, 武汉 430014)

摘要:通过对国内外农业信息化发展现状的研究,利用农业信息知识库理论基础,以农业产业链信息为核心,对农业政策、资讯、技术、标准等信息类别进行深入的研究,构建了以农业大数据为背景的开放式农业信息知识库,实现了农业资源整合、共享、分析及利用的功能,为帮助实现农业生产决策、农业病虫害的防治预警、农产品产量和价格预测等打下坚实的基础。

关键词:大数据;开放式;农业信息;知识库

中图分类号:G250.74;S126

文献标识码:A

文章编号:1003-8701(2018)05-0060-05

Studies on the Construction of Open Repository of Agricultural Information in Big Data Era

CHEN Pinting¹, GUAN Bo¹, SHEN Xiangcheng¹, WU Yaling¹, MA Hairong¹, DENG Dandan², ZHANG Yueting², XU Guimei², LUO Zhiqing^{1*}

(1. Institute of Agricultural Economic Technology, Hubei Academy of Agricultural Sciences, Wuhan 430063; 2. Wuhan Optical Valley Information Technology Co., Ltd., Wuhan 430014, China)

Abstract: Current situation of agricultural informatization at home and abroad was studied in the paper. Using agricultural information knowledge base, and taking the information of agricultural industrial chain as the core, agricultural policies, information technologies, standards and other information categories were deeply researched. The open agricultural information knowledge base in big data era was constructed, which fulfill the functions of agricultural resource integration, sharing, analysis and utilization. And it also laid a solid foundation for helping to achieve agricultural production decisions, prevention and warning of agricultural pests and diseases, agricultural product output and price forecasts, and so on.

Key words: Big data; Open; Agricultural information; Knowledge base

随着互联网的快速发展,农业网络数据呈爆炸式增长,但这些数据大部分是非结构化的,难以被计算机识别并利用。知识库是一类特殊的数据库,主要作用是领域知识的管理,包括知识采集、整理和提取等。由专家手工编撰的知识库虽然质量高,但是规模小,知识覆盖率低、更新缓慢,而且编撰过程需要耗费大量的人力和物力,

在大数据时代,如何应用网络优势构建一个大规模的农业知识库成为亟待解决的问题。维基百科、百度百科等均是以前信息资源共享为运行机制的网络百科全书,具有涵盖知识领域广泛、内容开放、允许大众参与的特点,是开放式知识库的典型案列。

近年来,国家对农业信息化高度关注,许多农业信息平台 and 网站应运而生,据统计,中国有近四万个农业网站,农业信息资源浩如烟海,但涉农企业和农户对农业信息的需求仍然有近90%得不到满足。农业科研工作和从事农业相关工作的人员无法及时全面精准地掌握最新的政策法规、研究成果、技术方法和发展趋势,大大降低了农业工作者的效率和能力的提升^[1]。随着“农业大数据”时代的到来,运用“互联网+农业知识服

收稿日期:2018-04-28

基金项目:湖北省农业科学院青年科学基金项目(2015NKYJJ28、2015NKYJJ34、2016NKYJJ34);湖北省农业科学院竞争性计划项目(2016JZXJH20、2015JZXJH08);湖北省农业科技创新中心(2016-620-000-001-042)

作者简介:陈娉婷(1982-),女,助理研究员,博士,主要从事农业信息化研究。

通讯作者:罗治情,男,博士,副研究员,E-mail: 32143018@qq.com

务”的模式,实现农业网络信息资源优化整合,开放共享等有效管理,是推动实现农业信息化、农业现代化的创新型模式。

1 国内外农业信息化建设研究

1.1 国际农业信息化建设研究

随着计算机技术、网络技术的快速发展及向农业的广泛渗透,在世界范围内掀起了农业信息革命的浪潮^[2]。目前,世界经济正在由工业化进入信息化,农业信息化已经成为现代农业的重要标志^[3]。经过近30年的农业信息化发展,信息技术在农业方面的应用已经由单纯的计算机统计和分析农业数据,到建立农业信息网络,对农业数据进行收集、传递、存储、分析、发布,构建大型农业数据库。在强大的数据背景支持下,开始实现对农业灾害的预警、远程专家诊断及自动化生产等方面的探索和实践。发达国家农业信息化已经进入了全面实现农业信息化时代,主要表现在网络和多媒体技术在农业中的应用及农业自动化控制技术在农业产前、产中、产后等全过程中的应用,向集成化、专业化、多媒体化和实用化发展的趋势。

美国建立较为健全的农业信息网,建成了 AGRICOLA、AGRIS、BIOSIS Preview、Pest Bank 等农业数据库对农业数据进行全面的采集整理和保存,这些数据库也实行完全开放制;此外,美国还拥有世界上最大的农业计算机网络系统 AGNET,该网络覆盖了美国的46个州,方便农户使用家里的电话或电脑就可以实现农业信息资源的分享,第一时间就可以了解农业市场信息。德国在联邦中央、州县和区实现了农业网络的全覆盖,建设有农业科学技术信息网络系统、食品和日用品检测方法数据库查询系统、农业科技期刊网络系统和数百个数据库,其中与农业相关的数据库就有30多个,为方便农户了解农业信息,获取服务提供了有力的保障。韩国建立农村振兴厅(RDA)数据库,聘请专业的工作人员对网上农业数据进行收集、加工,读者可以直接从网上订阅所需期刊和书籍,实现了农业信息资源及时、快捷的全民共享^[4]。印度开展的“邮车网络”和信息亭建设,为农民提供低价易用的计算机等,为印度的信息化领域打下了坚实基础。欧盟设有欧洲统计局、欧盟农业委员会、农场会计网委员会和农业理事会,为社会提供直接而广泛的农业信息宣传和职业培训^[5]。尽管各个国家的农业信息化发展

有所差异,但在农业信息化的高速发展时期,都离不开农业网络和农业数据库的开发利用。

1.2 国内农业信息化建设研究

我国农业信息化建设起步较晚,但近几年发展速度较快。从1986年的农业部《农牧渔业信息管理系统总体设计》到《农业部电子信息系统推广应用工作的‘八五’计划及十年设想》,再到《农村经济信息体系建设‘九五’计划和2010年规划》,我国农业信息化取得了一定程度的发展。自从十八大提出“四化同步”的战略部署以来,农业信息化建设得到了国家政府的高度重视和支持,各类涉农网站也如雨后春笋般层出不穷。据统计,国内的涉农网站已经增至近4万个,超过了英法等发达国家。同时,还建立了一系列具有代表性的数据库:中国农林文献数据库、中国农业文摘数据库、农副产品深加工题录数据库、植物检疫病虫害名录数据库、农牧渔业科技成果数据库、农产品集贸市场价格行情数据库等,其中一部分数据库已经达到国际领先的水平^[6]。此外,我国引进 AGRIS、IFIS、AGRISCOLA、CABI 这4大世界先进的农业数据库,极大地推动了我国农业信息化的进步。

1.3 存在问题

从国外农业信息化建设体系情况来看,在硬件上,发达国家已经基本实现了农业信息网络全覆盖的状况,在软件上,各类农业网络系统和各种农业信息数据库已经较为完善,并且传播和共享机制基本可以实现对农户发布及时、快捷、方便的农业信息,但由于发达国家的农业状况是以集约化、精准化农业为主要发展对象,而我国还处在传统的小农业模式向新型农业模式转型阶段,如果完全照搬国外农业信息化建设手段来达到发达国家的水平会有一些的难度。

与国外的农业信息产业相比,我国的农业信息化建设经过多年的努力,已在信息资源的开发和利用、自然资源数据的收集、农业信息管理和应用系统和数据库开发建设方面都获得较多的经验,在开展农业相关标准、规范的制(修)订方面,表现出与信息化建设的要求差距巨大。由于大量的农业信息都没有进行有效的分类、定义和描述,造成农业信息资源没有被有效地利用,同时也导致农业信息服务能力严重不足。

以知识库^[7]理论研究为基础,利用农业大数据和信息检索等计算机和网络技术,收集和整合涉农领域的政策、法规、知识、技术、专家等信息资

源,利用众源编纂的管理方式,搭建农业信息公共服务平台,构建以政策法规、行业资讯、农业知识、专家咨询、农产品市场信息、农业产业服务等为主要服务内容的一站式信息服务模式,是一种实现“互联网+农业大数据”,面向农业信息化领域的为用户提供及时、准确、便捷的信息传递方式,为实现我国农业信息服务的精准化、高效化提供了有利条件,为农业数据统计、规划、科技创新提供基础支撑作用。

2 知识库系统框架设计特点

农业信息知识库是为了满足大多数涉农企业和农户的需求而设计的,一方面要保持知识库系统具有稳定可靠的组织结构。另一方面,需要实现各类农业信息有效的集合,可以实现信息共享。因此,知识库的构建应具备易用性、实用性、可靠性、可延展性、开放性五大特点。

2.1 易用性

农业信息知识库的服务对象是科学文化水平不高的涉农企业和农户,要让他们能够在知识库平台上操作顺畅,简单易懂是构建本知识库的重要原则。农业信息知识库的易用性不仅表现在让用户操作起来简单,同时,在对农业知识的表达上尽可能用灵活多样的表现形式,比如:视频、照片等,让他们能更直观更深入地了解农业知识,达到能快速指导生产的目的。

2.2 实用性

农户最关心的问题是“种什么养什么”、“怎么种怎么养”以及“怎么卖”。那怎么才能让用户快速获取想要得到的信息,以及解决生产中遇到的技术问题?这是反映知识库实用性的关键。为使用者提供最新的农业资讯,最准确的农产品市场信息,最专业的种植、养殖的技术指导,才能为涉农企业和农户提供最大的帮助。

2.3 可靠性

农业信息知识库是一个专门服务于涉农主体的知识服务系统,是在各类农业信息数据库的基础上,对相关的信息知识资源按照一定的组织方式和结构整合起来,其中考虑到农业信息和知识的多元性,设计各种数据库、异构数据库等先进技术来保证知识库的系统正常稳定地运转。同时,由于多种数据库和信息资源结合的复杂性系统结构,开发资源整合技术来完成协调工作。

2.4 可延展性

作为一个可以长期使用并能及时更新的知识

库,必须具备可延展性。农业信息知识是在不断更新的,知识库的系统也需要与时俱进,当有更新的数据,有更先进的数据库技术时,旧系统应该具有与新系统顺利衔接的功能,既能保证旧的数据不丢失正常使用,又能和新数据再次整合,实现在新系统环境下,为涉农企业和农户提供更新更便捷的农业信息服务系统。

2.5 开放性

农业信息知识库的设计首先考虑众源编撰的模式,因为农业知识更新快,知识面广阔,很难由一个或几个专家负责收集、整理和编辑,应设计一个开放式的组织方式。同时,知识的表现形式也应多元化,不只拘泥于文字或图片,应增加多媒体数据资源的存储和表达,便于各个文化水平的人都能快速理解和掌握新的农业知识。

3 知识库系统框架和信息体系构建研究

3.1 知识库建设的基本思路

按照采用人工或自动数据采集来建立农业信息资源数据库的一般知识库的建设思路,难以满足涉农企业和农户对实时更新数据和为国家的农业决策分析服务的需求。本知识库建设的思路如图1所示,通过开放式的农业信息管理模式,围绕农业产业链形成的农业产业信息链条为使用者提供最及时、准确的农业信息,此外为农业专家提供一套知识库管理维护系统,可以随时对知识库进行新增、修改和删除等操作。通过农业信息知识库的应用,产生农业大数据,进而推动农业信息化和农业现代化的实现。

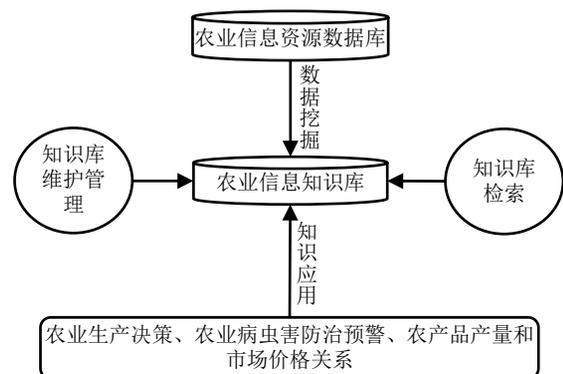


图1 农业信息知识库建设的基本思路

3.2 系统框架设计

农业信息知识库系统基于J2EE和云计算进行架构设计,采用异构SOA系统架构,具有非常强的适应性、扩展性和灵活性等。系统基于SaaS

的方式为用户提供服务,系统自身集成多种应用,并根据不同类型用户,通过不同终端应用提供定制化和个性化服务。同时,系统提供标准的API,通过接入第三方应用,丰富系统功能。在用户体验方面,采用当前主流的UI设计技术以及人

机交互方式,保证系统的可操作性、流畅性、友好性等,提升用户感知和用户体验。

3.3 信息体系构建研究

20世纪90年代以前,我国习惯性地将农业信息分为农、林、牧、渔四个类别。随着社会发

表1 知识库信息体系构建

| 信息分类 | 来源 | 数据类型 | 用途 | 主要服务对象 |
|------|--|--|---|---------------------------------|
| 政策法规 | 全球新兴市场商业资讯数据库EMIS、Factiva数据库、重庆维普中国科技经济新闻数据库、网络自动跟踪与自动采编信息等 | WORD PDF | 为政府和管理部门提供政策形成过程中的智力支持或科学依据;为涉农企业和农户个人明确农业生产方向、解读政策思想并提供合理的生产方案 | 政府与管理部门、科研院所和高校等农业科研人员、涉农企业、农户等 |
| 行业资讯 | 农副产品深加工题录数据库、植物检疫病虫害名录数据库、农牧渔业科技成果数据库、农产品集贸市场价格行情数据库、AGRIS、IFIS、AGRISCOLA、CABI等 | WORD PDF EXCEL | 提供行业内最新的各类农业咨询及动态,指导农业生产,公布农产品市场信息 | 政府与管理部门、科研院所和高校等农业科研人员、涉农企业、农户等 |
| 农业技术 | 科学引文索引SCI-EXPANDED、THOMSON INNOVATION、AGRIS、AGRISCOLA、CABI、中国科学引文数据库CSCD、万方数据库、清华同方中国期刊全文数据库、重庆维普中文科技期刊数据库、国外网络版外文期刊全文数据库、台湾学术文献数据库、中图在线、中国电子期刊等 | WORD PDF EXCEL AVI MPG RM RMVB WMV MP4 FLV MKV | 提供国内外期刊学术论文、会议论文、教学视频、电子书籍等科学技术信息的下载及推送等服务 | 政府与管理部门、科研院所和高校等农业科研人员、涉农企业、农户等 |
| 标准规范 | 世界标准检索数据库WSS、国外标准数据库、中国国家标准、行业标准、台湾标准目录数据库等 | PDF | 提供国内外标准的检索、下载、原文传递等信息服务,追踪标准的最新状态,如:标准是否被更新、替代、修订或作废等信息 | 农业科研工作人员、涉农企业和农户等 |
| 专利 | 德温特世界专利索引DWPI、美国专利全文数据库、世界知识产权数字图书馆、加拿大专利数据库、欧洲专利数据库、日本专利数据库、韩国知识产权、中国专利全文数据库、中国专利局专利检索、中国专利信息中心、台湾专利数据库等 | CSV EXCEL PDF TXT | 提供国外和中国国家知识产权局的检索和全文下载 | 农业科研工作人员、涉农企业和农户等 |
| 专家 | 湖北省农业专家大院数据库、国内外科研院所及高校的从事农业方面研究的专家数据库 | WORD PDF | 提供信息跟踪与互动,推动农业科技成果转化为市场应用新技术;实现提供与需求过程中农业技术传递的对接服务 | 涉农企业和农户等 |

农业信息发生爆炸式增长,简单的信息分类系统已经不能满足现有的信息体系建设。因此,笔者根据农业信息的来源、层次和类别,将农业信息资源按照一定的规则进行整合、分类,使其形成系统化、结构化的信息数据库(表1),使农业领域的各部门各行业的工作人员和涉农主体都能找到需要的信息,实现农业信息资源共享的目的(表2)。

表2 知识库服务体系

| 网站导航 | 情报咨询 | 用户服务 |
|----------|--------|------|
| 高清地图 | 政策解读 | 用户指南 |
| 数据库 | 公共查询 | 在线咨询 |
| 农产品名录 | 委托检索 | 用户评价 |
| 国家机关机构名录 | 农业产业分析 | 文件下载 |
| 涉农企业名录 | 行业热点推送 | 网上购物 |
| 专家名录 | 定制主题跟踪 | |
| 友情链接 | 专利分析 | |

4 农业信息知识库的应用

4.1 农业生产决策研究

建立农业信息知识库是农业生产决策研究的前提条件。随着农业知识库数据的不断积累,从大量的农业数据中提取潜在的有价值的知识是一件非常重要的研究。利用数据挖掘技术,对海量农业数据进行抽取和分析,找出数据与数据之间的发生关系和规律,掌握农业生产变化规律,构建基于农业知识库和大数据技术的生产决策系统;基于决策系统对综合数据的推理和分析,为政府提供可行的农业生产决策方案,为农业生产时间配置提供有效的指导和帮助。

4.2 农业病虫害的防治预警研究

农业病虫害直接影响农产品的产量和质量。因此农业病虫害预测对农业生产有着极其重要的意义。在采集全国的农业病虫害发生与防治数据

的基础上,对历年数据进行统计与分析,综合病虫害的种类、数量、程度等参数,建立病虫害发生的预警数据库,当某一些病虫害参数数值发生变化时,可能代表着病虫害灾害的发生。向农业生产者提供及时、准确的病虫害预警信息,可以有效地减少、防止更大病虫害灾情的发生,为政府部门农业预警提供可靠的数据支持。

4.3 农产品产量和 market 价格的规律研究

农产品的市场价格与农业生产者的利益息息相关,同时也是引发社会问题的关键因素之一。利用大数据分析农产品的需求量与 market 价格的内在联系,从而指导农民生产,稳定农产品价格,可以帮助实现国家粮食安全。例如,通过对农产品不同品种的生产者、批发商、零售商等历年销售价格数据的采集,同时对应的对当年的农产品产量数据进行采集,分析产量与价格数据之间的发生规律,对未来的农产品价格的预测,规划来年的农产品种植面积有着重要意义。

参考文献:

- [1] 钟丽丹,潘杏梅,王建涛.基于大数据开放式海洋信息知识库构建研究[J].图书馆杂志,2016,35(8):59-66.
- [2] 葛佳琨,刘淑霞.数字农业的发展现状及展望[J].东北农业科学,2017,42(3):58-62.
- [3] 曹俊杰.农业信息化建设的国际经验与中国模式[J].经济问题探索,2008(3):52-56.
- [4] 王丞,安森东,颜廷武.农业信息资源可持续开发利用的国际比较及启示[J].中国人口·资源与环境,2011,21(8):170-174.
- [5] 王玉芹,杨晓蓉.韩国农业信息技术的特点和发展方向[J].情报杂志,2004(10):95-99.
- [6] 吕晓燕,卢向峰,郝建胜.国内外农业信息化现状[J].农业图书情报学刊,2004,16(11):121-125.
- [7] 董文鸳,袁顺波.全球学科知识库发展现状扫描[J].图书馆,2015(4):40-43.

(责任编辑:王昱)