

基于语义技术的柑橘园土壤环境判定决策支持系统的设计与实现

朱 颖 张自力

(西南大学计算机与信息科学学院 重庆 400715)

摘 要 农业生产管理决策支持系统对提高相关农产品的产量与质量起到越来越重要的作用。针对柑橘生产中土壤环境影响柑橘生长的问题,提出了基于语义技术的柑橘园土壤环境判定决策支持系统,重点讨论了该决策支持系统的系统结构、土壤语义数据库的建立以及推理规则的定义等,并应用语义数据库软件 AllegroGraph 实现了柑橘土壤语义数据库。

关键词 决策支持,语义技术,本体,柑橘,土壤

中图分类号 TP182

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2014.09.055

Design and Implementation of Citrus Orchard Soil Environmental Decision-making Support System Based on Semantic Web

ZHU Ying ZHANG Zi-li

(College of Computer and Information Science, Southwest University, Chongqing 400715, China)

Abstract Agricultural production management decision-making system plays more and more important role in improving the yield and quality of agricultural products. In this paper, according to the soil environment problems in citrus production, we proposed an approach based on Semantic Web to build citrus orchard soil environmental decision-making support system. We focused on the system structure of the decision support system, the establishment of soil semantic database and definition of reasoning rules. Finally, we applied the semantic database software AllegroGraph to realize the citrus soil semantic database.

Keywords Decision-making, Semantic technology, Ontology, Citrus, Soil

1 引言

柑橘是世界第一大类水果,其种植面积与产量均居首位^[1]。柑橘生长的土壤是由多层矿物质组成的自然体,是其生产的物质基础。土壤的理化性质与土壤中含有的营养元素的丰缺,是制定柑橘园土壤管理和果树合理施肥方案的重要依据之一,对柑橘的正常发育及果品质量与产量具有重要作用,甚至在一定条件下起决定性作用^[2]。

在目前的柑橘生产中,要实现果园养分高效、作物高产,同时保护果园自然环境,了解土壤的养分状况是前提条件,科学施肥是技术手段。而我国柑橘园在柑橘生产过程中管理粗放,果农在不了解橘园土壤环境的况下,凭经验盲目施肥,造成了肥料利用率低、肥料施用成本高、橘园单产不高、果品不佳的状况。

随着信息技术的迅猛发展,为了解决农业生产中所存在的上述问题,以物联网为基础的精准农业生产管理系统成为研究热点,旨在为农户提供农业生产相关的信息服务。柑橘土壤环境判定决策支持系统作为柑橘精准生产管理系统的一部分,是以物联网为基础的,其核心为关于橘园土壤环境判定

的智能决策系统。目前绝大多数与农业相关的决策支持系统仍然使用关系数据库,这使得在处理文本型的农业生产知识时限制较多,不够灵活,从而不能为农户提供个性化的主动的信息,且不能自动推送农户在农业生产不同阶段所需的信息。因此,我们将语义技术应用到智能决策系统中,利用语义技术对海量、异构、多源的农业生产管理知识进行语义集成。在该系统中,语义技术的应用主要体现在以下两方面:1)通过构建本体生成语义数据库,与传统的关系数据库相比,语义数据库的特点是数据的添加、更新、删除更加灵活,查询更加精确,便于处理多维度问题;2)在语义数据库之上的语义推理,语义推理是利用某种策略从已知推断出另一种判断的过程^[3]。

本文第2节主要介绍了语义技术和本体的相关理论;第3节给出了柑橘土壤环境判定决策支持系统的体系结构;第4节为柑橘土壤语义数据库的建立;最后总结全文。

2 语义技术与本体

2.1 语义技术

语义万维网(Semantic Web)的概念自1999年由Tim

到稿日期:2013-11-18 返修日期:2014-02-16

朱颖(1982—),女,硕士生,主要研究方向为数据库、知识库的应用,E-mail:zhuyier420@126.com;张自力(1964—),男,教授,博士生导师,主要研究方向为智能系统的理论与实现技术。

Berners Lee 等人提出以来,短短十余年间,吸引了各领域学科的众多学者进行科学研究,语义万维网逐渐成为下一代万维网的重要发展方向^[4]。语义网的目标是使得网络上的信息具有计算机可以理解的语义,从而为人们提供各种智能服务。在语义网中,XML、RDF 和 Ontology 是核心内容^[5]。RDF 解决如何无二义性地描述资源对象的问题,使得描述的资源的元数据信息成为机器可以理解的信息。RDF 通过基于 XML 语法的明确定义的模型,来建立语义协定(RDFS)和语法编码(XML)之间的桥梁,并以此来实现元数据的互操作能力。本体层是在 RDF(S)基础上定义的概念及其关系的抽象描述,用于描述应用领域的知识,描述各类资源及资源之间的关系。

2.2 本体

17 世纪,德国经院学者郭克兰纽(Goclenius)首先提出并使用了本体论(Ontology)一词。近几十年里,这个词被应用到计算机领域,并在人工智能、计算机语言以及数据库理论中扮演着越来越重要的角色。1993 年 Gruber 提出“本体是概念化的明确的规范说明”的定义,该定义被广泛引用^[6]。1998 年,德国学者 Studer 等人在 Gruber、Borst 的基础上,给出了“本体是共享概念模型的明确的形式化说明”^[7]。Neches 等人将本体定义为“给出构成相关领域词汇的基本术语和关系,以及利用这些术语和关系构成的规定这些词汇外延的规则的定义”^[8]。

本体通过对特定领域知识、领域专家意见及权威参考文献进行概念组织,明确该领域术语及术语间的关系定义,从而达到对该领域知识的共同认识与理解。总的来讲,本体是一种计算机能够理解的、结构化的语义描述机制,是使某领域内的概念及概念间的关系得到规范化和形式化的描述,以达到知识共享的目的,实现系统之间信息的共享与交互、信息处理的自动化。

2.3 RDF/RDFS 描述资源框架

资源描述框架 RDF (Resource Description Framework, RDF)^[8]是一种通用的元数据结构,建立在一个称为“三元组”的数学模型上,它可以把一些非常简单的元数据说明组合在一起。每一个“三元组”由主语(Subject)、谓语(Predicate)和宾语(Object)这 3 个元素构成,三者组成一个语句(Statement),多个语句的集合称为描述(Description)。主语代表资源,即所有可由 RDF 表达的对象;谓语是资源的属性;宾语是属性的值,每个属性值可以是简单的值,如数字、字符串等,也可以更复杂,这些值本身又是资源。

RDF 具有如下特点:1)易控制。因为其使用简单的资源-属性-值三元组。2)易扩展。因为词汇集和资源描述是分开的。3)包容性。允许任何人定义自己的词汇集,并可以根据需要,无缝地使用多种词汇集来描述资源。4)可交换性。使用 XML 语法,可以很容易地在网络上实现数据交换。5)易综合。在 RDF 中资源的属性是资源,属性值可以是资源,关于资源的陈述也可以是资源,都可以用 RDF 来描述,这样就可以很容易地将多个描述综合,以达到发现知识的目的。

RDFS 是在 RDF 的基础上定义的一套简单的模式语言。RDFS 提供了一些建模原语,用来定义一个描述类、类与类之间关系的简单模型,这相当于为描述网上资源的 RDF 语句提供了一个词汇表,从而增强了 RDF 对资源的描述能力。

2.4 语义推理

在传统关系数据库中,数据之间的逻辑联系没有被显式

描述,而是隐藏在数据库表之间的联系、文档和 SQL 代码中,需要使用查询语言来获取,导致计算机程序无法准确地找出和运用这些隐藏的逻辑关系和信息。相反,RDF 语义数据库中知识个体之间建立了明确的逻辑联系,在其上进行的语义推理是一个自然的、简单的过程。推理规则由农学专家提供,并通过图形化操作界面、软件平台(工作台)进行添加、删除和修改。新规则输入之后,需要进行一致性检查,以确保新输入规则与已有规则满足统一性和一致性的要求。这些规则被自动转换为程序语言,例如 Prolog 或者 Jess,系统可以编译并使用。

3 系统框架

基于物联网的柑橘土壤环境判定决策支持系统(见图 1)是利用语义技术,将已有柑橘种植土壤知识、养分及肥料知识等进行语义集成,形成柑橘土壤知识本体和语义数据库。在 RDF 语义数据库的基础上,构建推理规则,进行自动语义推理,结合由物联网终端采集到的柑橘园土壤环境因子包括:土壤水分、pH 值、土壤养分等,为果农自动提供个性化的决策服务,帮助用户利用丰富的专家知识来更好地进行农业生产。农户可以通过现有通讯设施,如移动互联网、3G、无线网络等,以手机短信、智能手机应用等方式接受系统推送的实时信息。该系统由 4 部分组成。

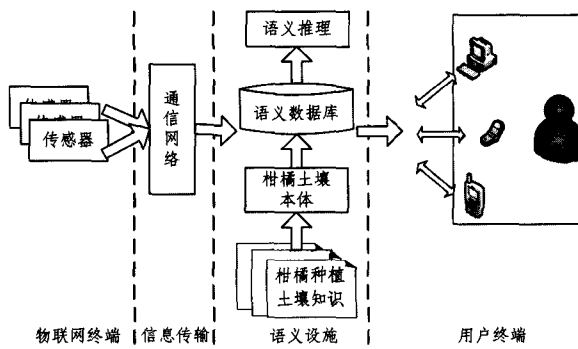


图1 柑橘土壤环境判定决策支持系统

①物联网终端:利用基于无线传感器的信息采集网络,采集橘园土壤环境信息,并将信息发送给远程数据服务器,由数据服务器进行下一步的分析与处理,实现果园的土壤环境的远程监测。例如,2008 年,中国农业大学精细农业研究中心成功研制了太阳能供电土壤湿度无线传感器网络设备,国家农业信息化工程技术研究中心成功研制了电池供电的田间环境/土壤信息获取无线传感器网络设备。彭燕、何东健提出了一种基于 Zigbee 技术的果园生态环境监测系统,该方案基于无线通信协议 Zigbee 技术,采用 Microchip 公司的 Zigbee 协议栈和开发平台,构建了星型无线传感器网络。

②信息传输:位于柑橘园中的物联网无线传感器所采集到的果园土壤环境数据,通过某个私有信道进行传输。该私有信道可由当地移动通信运营商建立,而其他信息的传输,可通过基础网络设施来进行。

③语义设施:语义设施在柑橘土壤环境判定决策支持系统中处于核心地位。语义技术是实现该系统的关键技术。将柑橘种植土壤知识转换为 RDF 三元组,RDF 的优点在于简单、开放、易扩展、易修改、易交换。通过转换,我们构建出柑橘种植土壤知识本体,本体是柑橘种植土壤知识形式化与概

念化的规范说明,是对柑橘土壤知识的一致性描述,是建立推理规则,进行自动推理的基础。而构建柑橘种植土壤本体,生成语义数据库,是整个决策支持系统的重要基石。决策支持系统将根据语义数据库中的土壤种植知识,结合由物联网传感器所采集到的柑橘园土壤环境数据,为用户提供主动的、个性化的信息服务。

④用户终端:通过互联网或者移动通信网络,用户可通过用户终端向柑橘园土壤环境判定决策支持系统发送、接收信息。用户终端可以是个人电脑、手机或智能手机等。而决策系统不仅在用户提出请求时给予回馈,而且在一定条件下能够主动将橘园土壤环境信息推送给用户。

4 柑橘种植土壤语义知识数据库的建立

柑橘土壤环境判定决策支持系统的核心为语义设施,因此我们的首要工作是构建柑橘土壤知识本体,建立柑橘种植土壤知识相关的语义知识数据库。本文以重庆三峡库区柑橘园为背景,构建了重庆三峡库区柑橘种植土壤知识库。在构建柑橘种植土壤本体时,我们使用了本体建模工具 TopBraid Composer(TBC)^[10]。TBC 基于 Eclipse 开发而成,是一个企业级的语义网本体建模和语义应用开发工具。该工具完全符合 W3C 标准,是目前处于领先水平的 RDF 与 OWL 本体编辑器,同时内置推理引擎、SWRL 编辑器、SPARQL 查询工具等。AllegroGraph 是由美国 Franz 公司出品的一个基于 W3C 标准的为资源描述框架构建的图形数据库^[11]。Allegrograph 可以实现 10 亿级别的三元组数据库查询,提供了完整的对 RDF 数据的事务管理(ACID)的支持,像关系数据库一样支持 RDF 数据事务的回滚、提交和存储过程。Gruff 是一款强大的可视化 RDF 浏览器,用于改善数据检索的过程,可以对本地以及网络的 RDF 数据进行可视化操作。

4.1 柑橘种植土壤知识提取

在柑橘种植过程中,土壤部分的专业知识大部分以报告、文档等非结构化的形式存在,无法被计算机程序直接用来指导农户进行生产管理。构建柑橘种植土壤知识库的第一步,就是利用语义技术通过人工及计算机程序从现有土壤信息数据中提取多年累积的柑橘生产种植土壤知识,尽量提取出一段话中的主语、谓语和宾语。知识的预处理要对概念事实标准化,不能有模糊不清的概念。例如土壤 pH 值的描述为“pH<4.5 为强酸性,4.5~5.4 为酸性,5.5~6.5 为酸性至弱酸性,6.5~7.5 为弱酸至弱碱性,7.5~8.5 为碱性,pH>8.5 为强碱性。在土壤 pH 值为 4.5~8 时,柑橘能够生长,最适宜柑橘生长的土壤 pH 值为 5.5~6.5,即弱酸性土壤^[12]”。我们对其进行预处理,得到土壤 pH 值分类,见表 1。

表 1 土壤 pH 值状况分类表

土壤	强酸性	酸性	酸性至弱酸性	弱酸至弱碱性	碱性	强碱性
pH 值	<4.5	4.5~5.4	5.5~6.5	6.5~7.5	7.5~8.5	>8.5

4.2 柑橘种植土壤本体的构建

类是语义本体中表示对象的概念集合,同时也是知识的集合,是语义本体的基础。在本体构建的过程中,除了定义类,还应定义类的属性,它用来描述已经定义的类之间的内部结构。例如,根据在上一步提取出的柑橘土壤 pH 值知识,我们在 TBC 中建立一个名为“土壤”的类。在“土壤”类下,我们

还建立了 3 个属性,分别是“Quantity Max”、“Quantity Min”和“Soil pH Value”。在定义属性之后,我们还应定义属性限制,即定义属性的取值范围、取值类型等相关特征。比如“土壤”类属性“Quantity Max”、“Quantity Min”和“Soil pH Value”所对应的取值类型应该是数据类型 xsd:float。

我们在“土壤”类的下面建立了 6 个实例,分别是“强酸性土壤”、“酸性土壤”、“酸性至弱酸性土壤”、“弱酸至弱碱性土壤”、“碱性土壤”以及“强碱性土壤”。图 2 为在 TBC 中建立的类与实例界面图。在 TBC 中我们先后建立了上面的类、属性、实例,确定了实例属性的取值,在建立的过程中实例作为主语(subject)如“酸性土壤”,属性作为谓语(predicate)如“Quantity Max”,属性值作为宾语(object)如“5.4”,自此新的一个 RDF 主(subject)-谓(verb)-宾(object)三元组被建立。图 3 为利用 Gruff 软件根据 TBC 中所构建的三元组所建立的“土壤”RDF 拓扑逻辑结构图。

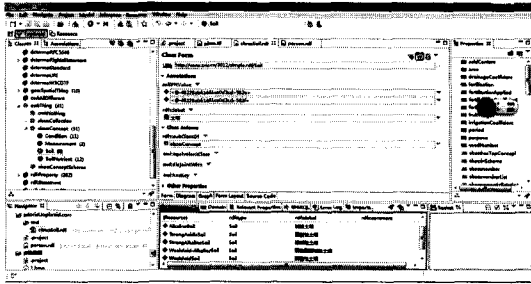


图 2 TopBraid Composer 界面

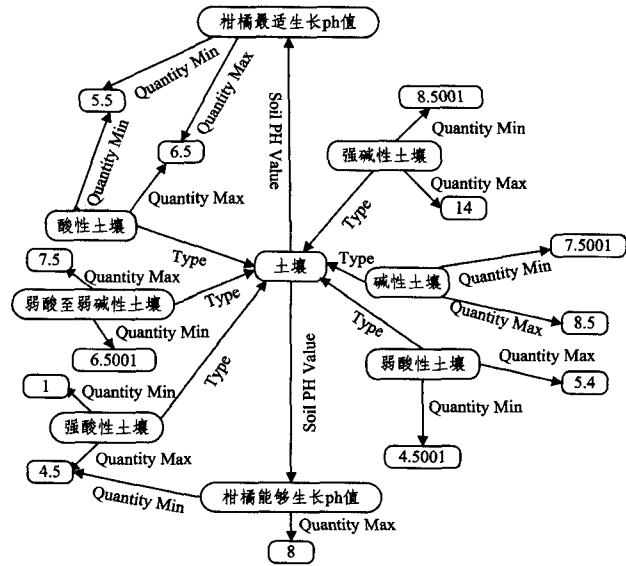


图 3 土壤 pH 值 RDF 三元组拓扑逻辑

4.3 柑橘种植土壤语义数据库

根据上述步骤,我们利用 TBC 将柑橘种植土壤知识构建为柑橘种植土壤本体,并将语义本体上传到 AllegroGraph 服务器中。建立柑橘种植土壤知识语义数据库是整个项目的核心工作。在语义数据库的基础上,系统能根据农户提出的服务请求,启动相应的语义推理获得决策支持,将土壤诊断结果返回给农户。系统也可以根据由物联网终端的无线传感器传回的数据结合数据库中的土壤知识,发现柑橘园土壤的潜在问题,主动向农户发送信息,例如土壤中某养分含量偏低,应采取何种措施等。

4.4 Prolog 规则

规则在决策支持系统中作为专家知识的一部分,也是语义技术的重要组成部分。在柑橘种植土壤知识中,也有一部分规则,如根据土壤中有机质含量的高低,判断土壤有机质对于柑橘生长是否适宜。我们采用 Prolog 语言来编写这些规则。

有机质含量极低:

```
(←--(low_organic_soil?soil)
(q-?soil !ex:organic_content ?content)
(lisp(<=(upi->value ?content) 5)))
```

有机质含量低:

```
(←--(low_organic_soil?soil)
(q- ?soil !ex:organic_content ?content)
(lisp(> (upi->value ?content) 5))
(lisp(<=(upi->value ?content) 10)))
```

有机质含量偏低:

```
(←--(low_organic_soil?soil)
(q- ?soil !ex:organic_content ?content)
(lisp(> (upi->value ?content) 10))
(lisp(<=(upi->value ?content) 15)))
```

有机质含量适宜:

```
(←--(low_organic_soil?soil)
(q- ?soil !ex:organic_content ?content)
(lisp(> (upi->value ?content) 15))
(lisp(<=(upi->value ?content) 30)))
```

有机质含量丰富:

```
(←--(low_organic_soil?soil)
(q- ?soil !ex:organic_content ?content)
(lisp(> (upi->value ?content) 30)))
```

4.5 SPARQL 查询

在完成上述步骤之后,我们可以对生成的模型进行测试,以检测模型是否达到预期目的。SPARQL^[13]查询语言是访问 RDF 数据的标准。SPARQL 查询语言与 RDF 紧密相连,不论哪种 RDF 查询图,都能利用 SPARQL 得到相应的查询结果。例如“酸性土壤的 pH 值是多少?”,利用 SPARQL 进行查询,图 4 为查询过程,其中左边为查询语句,右边为查询结果。

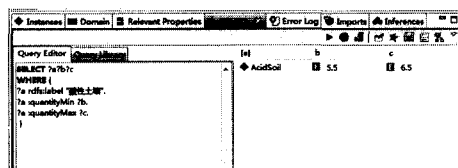


图 4 SPARQL 查询界面

总之,柑橘土壤环境判定决策支持系统建立在 Allegro-Graph 服务器的基础上,即系统的核心为语义数据库。系统根据语义数据库中的相关知识,可为农户提供主动及个性化的信息服务。例如,当农户通过智能手机或个人电脑发起一个查询请求“酸性土壤的 PH 值为多少”时,系统将利用

SPARQL 查询,查询出酸性土壤的 pH 值为 5.5~6.5,将结果返回给用户,同时提醒用户 pH 值为 5.5~6.5 适宜柑橘生长,农户不需采取任何措施。同时,系统可根据柑橘园中物联网传感器传回的土壤环境信息,如当前柑橘园某区域土壤的 pH 值,判断该 pH 是否在正常范围(适宜柑橘生长),并将结果主动返回给农户,将土壤管理意见发送至农户手机或个人电脑。

结束语 农业信息化是国家信息化建设的重要组成部分,随着现代化农业的发展,如何利用现有新兴技术,实现柑橘生产管理的智能决策,提升柑橘生产的科技水平,增加果农与企业的收入,是柑橘生产信息化的迫切需求。本文从柑橘生产土壤管理出发,区别于传统专家知识系统,结合语义技术,提出了基于语义的柑橘园土壤环境判定决策支持系统,旨在为农户提供主动的、个性化的、准确的信息。本文重点讨论了系统中语义数据库的构建,而由于时间、条件的限制,该系统还有待进一步完善。

参 考 文 献

- [1] Alva A K, Mattos Jr D, Paravasivm S, et al. Potassium Management for Optimizing Citrus Production and Quality [J]. International of Fruit Science, 2006, 6(1): 3-43
- [2] 王秀英, 石孝均. 重庆地区柑橘园土壤养分现状 & 优化施肥研究 [D]. 重庆: 西南大学, 2011
- [3] Blanco R, Halpin H, Herzig D M, et al. Repeatable and reliable search system evaluation using crowdsourcing [C]// Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2011: 923-932
- [4] 李洁, 丁颖. 语义网关键技术概述 [J]. 计算机工程与设计, 2007, 28(8): 1831-1833
- [5] Allemang D, Hendler J. Semantic Web for Working Ontologist: Effective Modeling in RDFS and OWL [M]. Elsevier, 2001
- [6] 王英林, 张申生. 基于本体映射规则的软件集成重构研究 [J]. 计算机学报, 2001, 24(7): 776-783
- [7] Studer R, Benjaminsand V R, Fensel D. Knowledge engineering: Principles and methods [J]. Data & knowledge Engineering, 1998, 25(1): 161-197
- [8] Neehes R, Fikes R, Finin T, et al. Enabling Technology for Knowledge Sharing [J]. AI Magazine, 1991, 12(3): 36-56
- [9] Lassila O, Swick R. Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation, World Wide Web Consortium [EB/OL]. (2004-02-10) [2008-07-15]. <http://www.w3.org/TR/REC-rdf-syntax/>
- [10] Weiten M. Ontostudio® as an ontology engineering environment [M]. Berlin: Springer, 2009: 51-60
- [11] AllegroGraph 4.3 [EB/OL]. <http://franz.com/agraph/allegro-graph>
- [12] 淳常品, 彭良志, 江才伦, 等. 三峡库区部分柑桔园土壤营养状况的初步研究 [J]. 中国南方果树, 2009, 38(2): 1-6
- [13] Hommeaux E P, Seabrne A. SPARQL query language for RDF [R]. W3C Recommendation, 2008, 15