

《智能信息处理》课程考试

国内本体研究进展

刘丙富

考核	到课[10]	作业[20]	考试[70]	课程成绩[100]
得分				

2020 年 11 月 27 日

国内本体研究进展

刘丙富

摘要 作为知识层面描述信息系统的概念建模工具和语义网核心技术，本体技术通过提供共同词汇解决人机信息交换问题、借助本体映射使不同模型方法、软件工具、范例与语言实现互操作、用计算机可理解形式化语言进行知识表示、获取、需求分析等，并广泛用于人工智能、知识工程、数字图书馆、信息检索、电子政务等领域。目前，本体技术研究已趋成熟并逐步实现工程应用但仍需深入研究：本体建模元语、构建标准等概念性问题达成共识，但本体半自动构建方法与工具、本体评价等尚未形成统一标准。客观、全面分析国内本体技术研究进展，对把握研究主题与热点、突破难点、促进技术发展有重要意义。

关键词 概念建模，语义网，本体，本体构建

Research progress of Ontology

Liu Bingfu

Abstract As a conceptual modeling tool and semantic web core technology for describing information system at knowledge level, ontology technology solves the problem of human-computer information exchange by providing common vocabulary, makes different model methods, software tools, examples and languages interoperate with ontology mapping, and uses computer understandable formal language for knowledge representation, acquisition, requirement analysis, etc., and is widely used in manual work Intelligence, knowledge engineering, digital library, information retrieval, e-government and other fields. At present, the research of ontology technology has become mature and gradually realize engineering application, but it still needs to be further studied: the conceptual issues such as ontology modeling meta language and construction standards have reached a consensus, but the semi-automatic ontology construction methods and tools, ontology evaluation have not yet formed a unified standard. Objective and comprehensive analysis of the research progress of ontology technology in China is of great significance to grasp the research topics and hot spots, break through the difficulties and promote the development of technology.

Keywords Concept modeling, Semantic web, Ontology, Ontology construction

1 引言

本体原本是一个哲学概念。在哲学意义上，本体论是对客观存在的一种系统的解释或解释，涉及客观现实的抽象本质。随着人工智能的发展，人工智能给出了本体的新定义。然而在开始的时候，人们对本体论的理解并不完善，这个词的定义是不断演变的。最广泛接受的本体定义之一：本体是共享概念模型的明确的形式化规范说明。

本体论反映的是被所有人接受的共同知识，反映的是相关领域的共同知识。本体的目标是捕获相关领域的知识，提供对领域知识的共同理解，并确定领域内共同识别的概念是由不同层次的形式化模型给出的，这些概念(术语)以及它们之间关系的明确定义。

因此，它是解决信息和知识共享与交换的语义层次的基础，具有良好的概念层次和逻辑推理的基础，已经广泛用于信息检索中，尤其是在基于概念的检索中。传统的论文数据库检索系统采用基于词匹配

的全文和标题关键词检索，往往检查的不够全面，也不够准确，会出现效果不好的情况。但是在设计中加入本体的话，不论是在知识层面，还是在概念层面上，都相当于增加了相关方面的知识体系，能够使用户检索到的信息更加准确、有效。

2 本体理论研究

国内本体技术研究吸收国外研究成果，重视理论研究，主要包括本体技术演变、定义、分类、描述语言、构建工具等。

2.1 本体概念

不同领域对本体定义不同。人工智能领域，Neches 等认为本体是给出构成相关领域词汇的基本术语和关系及用这些术语和关系构成的规定词汇外延规则的定义；B.Chandrasekaran 等认为本体研究特定领域知识的对象分类、对象属性和对象间关系，为领域知识描述提供术语。信息科学和知识工程领域，Gruber 认为本体是概念模型的明确的规范说明；Borst 认为是共享概念模型的形式化规范说明；Guarino 等引入“域空间”概念并定义“概念关系”，明确说明本体论和概念化间差别，对“概念化”明确说明；Studer 等认为是共享概念模型的明确的形式化规范说明；Fonseca 从建模元语方面认为是从特定角度用特定词汇表描述实体、概念、特性和相关功能的理论；William 等从特征和形态方面认为是用于描述或表达领域知识的概念或术语集，以组织知识库较高层次抽象或描述特定领域知识。总之，本体主要包含概念模型、明确、形式化和共享四层含义。

2.2 本体分类

本体按不同标准（如表示形式化度、描述对象、应用领域、目的、主题、是否在线、是否共享等）分类方式各异：Guarino 基于详细度（分为详细度高的参考本体及详细度低的共享本体）和领域依赖度（分为顶级本体、领域本体、任务本体和应用本体）分类；按本体形式化程度分高度非形式化本体、非形式化本体、半形式化本体、严格形式化本体；按描述对象

分特殊领域本体、一般世界知识本体、问题求解本体和知识表示语言本体等；按应用领域分为人或组织的通信、系统间互操作和系统工程领域三类。此外，Perez 和 Benjamins 将本体分为知识表示本体、普通本体、顶级本体、元（核心）本体、领域本体、语言本体、任务本体、领域-任务本体、方法本体和应用本体 10 种有交叉本体。

2.3 描述语言

本体描述语言应有定义良好的语法、语义、有效推理支持、充分表达能力，以清晰、形式化描述概念，可用自然语言、框架、逻辑语言、语义网络等描述。本体描述语言根据应用及功能各异：基于 AI 本体描述语言包括 KIF（多用于企业级）、Ontolingua、Loom、OCML、OKBC、FLogic 等；基于 Web 本体描述语言包括 SHOE（基于并扩展 HTML）、OML、CMKL、OWL（包括 3 个子语言 OWL Lite、OWL DL、OWL Full）、XOL、DAML、OIL、RDF、RDF(S)（后 6 种均基于 XML，属 W3C 本体语言栈中不同层次）等，此外，OWL、OIL、DAML+OIL、CMKL 基于描述逻辑，OWL、DAML+OIL、RDF(S) 是 W3C 推荐的标准本体描述语言；基于谓词逻辑本体描述语言包括 Ontolingua、CycL、Loom、OIL、KIF、F-logic、XOL 等，前三者和具体系统相关，基本只用于相关项目；基于图本体描述语言包括 WordNet 语义网络、概念图、CR、DAG、LSG、LCG 等。

2.4 构建工具

按支持语言分两类：基于特定语言并一定程度支持多种基于 AI 本体描述语言，如基于 Ontolingua 语言的 Ontolingua、基于 LOOM 语言的 OntoSau-rus、基于 OCML 语言的 WebOnto 等；基于 Web 本体描述语言构建语义 Web 本体，如 Protégé、WebODE、OntoEdit、OilEd 等。按功能分编辑工具（如 On-toEdit、OilEd、Protégé 等）、标注工具（如 AeroDAML、OntoAnnotate、COHSE 等）和集成工具

(如 PROMPT、FCA-Merge、ODEMerge 等)、本体存储查询工具和学习工具等。

3 本体构建与优化

3.1 构建方法

本体构建遵循明确性和客观性、完全性、一致性、最大单调可扩展性、最小承诺等规则,主要构建方法有七步法、MEHNLOGY 法、IDEF5 法、TOVE 法(又称 Gruninger&Fox “评价法”)、骨架法(又称 En-erriseontology 法)、SENSUS 法、KACTUS 法(又称 Bernerasetal 法)、AFM 法、五步循环法、循环获取法、生命周期法和基于领域知识重用的虚拟本体构造法等。

3.2 建模元语

主要包含类(又称概念)、关系、函数、公理和实例 5 类,概念间基本关系包括 part-of、kind-of、in-stance-of 和 attribute-of。实际应用中本体构造无需严格按上述 5 类元语且概念间关系定义可按应用需求而不限于上述 4 种基本关系。

3.3 建模思路

核心包括明确领域中概念、概念属性和约束条件、概念间层次关系等。陈刚等提出重用现有领域知识库知识构建可重用虚拟领域本体思路;王洪伟等提出基于描述逻辑的本体建模思路并构建基于逻辑的形式化本体模型;向阳等根据软件工程原理,提出基于 Jena 的 OWL 本体建模思路,包括构建步骤与实现算法;刘宗田等提出构建面向事件的本体模型;丁晟春基于本体工程思想提出基于顶层本体的领域本体构建框架。

3.4 本体映射

旨在发现不同领域本体概念间相关性,是本体集成、合并、修正、翻译技术基础,可很好解决本体异构问题,主要研究本体映射方法、工具、模型等。黄烟波等分类本体映射方法,针对本体定义模型提出基于语法、实例、定义和层次结构的分类,针对映射技术提出基于规则、统计学和机器学习的分类;瞿裕忠等从模型转换途径、映射策略适用范围及映射结果表

达形式三方面归纳关系数据库模式与本体映射难点并基于此比较映射工具;唐杰等基于贝叶斯决策理论提出最小风险的本体映射模型 RiMOM,提供多策略本体映射方法;高鹏等研究用户需求本体与产品配置本体间映射,以自动获取产品配置模型;徐振宁等提出本体到文档类型定义映射算法和文档类型定义到关系视图映射算法,实现 XML 文档和数据库视图在语义和结构上对应关系。王宇华等构建本体需求模型到 UML 原模型映射规则以实现两者转换;刘宗田等提出用事件本体将句子或篇章映射成事件网络。

4 本体应用研究

4.1 语义 Web

本体技术对语义 Web 研究、应用起重要推动作用。李曼等用领域本体及其推理能力生成优化的服务组合图,提出 Web 服务动态组合方法;吴健等基于词语间距离度量、相似度两种词汇语义相似度算法,提出基于本体论和词汇语义相似度的 Web 服务发现方法;彭晖等提出基于本体概念相似度(通过概念间语义距离计算)描述服务请求方和发布方的 Web 服务匹配算法;邓志鸿等分析本体在 Web 信息集成中应用;周明建等提出基于 Web 页面的信息项本体和结构项本体的信息提取规则以有效提取 Web 页面信息;袁柳等提出基于领域本体语义标注 Web 数据库查询结果以便机器处理和用户理解。

4.2 信息检索

本体技术可有效提高信息检索系统性。廖明宏提出基于本体信息检索方法;徐振宁提出基于本体的智能化、个性化语义信息检索系统体系结构;武成岗等提出基于本体和多智能主体、对用户需求及 Internet 信息进行领域分类的信息检索服务框架以提高检索结果针对性;万捷等基于本体将用户检索需求扩充成语义集并通过文档分析器过滤检索结果以提高检索质量;丁晟春等分析 Jena 在语义检索中作用与应用;廖乐健等从知识表示与推理角度提出基于本体与模板规则混合技术,混合

本体和树形模板以增强模板语义表达能力,提高信息抽取智能性。

4.3 知识工程

顾芳概述了知识工程中本体构建准则、方法、表示语言、代表项目、主要应用等;王英林提出基于本体重构知识管理系统框架,解决知识管理中知识类型不可扩充局限性;郭鸣提出基于本体和语义 Web、支持知识处理的结构层次化产品信息模型并给出从 EXPRESS 模式到 DAML+OIL 映射方法;胡玉杰等基于产品知识表达模型构建流本体和功能本体,进而定义特定领域产品共享、通用知识并提出基于本体的产品知识表达应用模型和集成框架。

4.4 其他研究

主要是本体在关系数据库与本体库中不同存储格式间转换问题。李曼提出将常用本体查询信息按类分别存于不同表以减少本体查询时表连接代价;徐振宁将知识表示和处理引入到 Web 信息处理,为半结构化 Web 数据和关系数据库提供统一语义模型,实现基于数据库的 Web 信息动态发布与多数据源集成。

5 小结

本体的构建过程离不开领域专家的参与。从知识基础角度看,主要分发展演绎、理论研究、本体构建与优化、本体应用四方面:发展演绎表明本体技术研究趋向多元化发展、交叉学科应用;理论研究较多,虽尚无完全统一标准但基本达成共识,包括本体定义、分类、描述语言、构建工具等,学者主要分布于人工智能、信息科学领域;本体构建与优化涉及构建方法、建模思路、本体映射、相似度计算、本体评价等,本体优化、本体评价、本体半自动构建等将逐步成为研究热点;本体主要用于语义 Web、信息检索、需求分析、知识工程等领域。

参考文献

- [1]孙雨生,付荣荣,郭隆敏.国内本体研究进展:载文分析和知识基础[J].计算机与数字工程,2020,48(06):1314-1323+1378.
- [2]姜玉红,鲍玉来.基于本体的领域概念语义描述研究——以中华武术术语为例[J].情报科学,2020,38(08):141-144+169.
- [3]杨利,叶明全,郑诚.基于领域本体概念划分的语义标注方法[J].合肥学院学报(自然科学版),2013,23(01):41-44.
- [4]裴培,丁雪晶.基于本体的语义相似度计算综述[J].合肥学院学报(综合版),2020,37(05):68-74.
- [5]张克亮,李芊芊.基于本体的语义相似度计算研究[J].郑州大学学报(理学版),2019,51(02):52-59.
- [6]李林.基于领域本体和知识图谱的标准符合性测试[J].科技风,2020(32):100-101.
- [7]方朝晖.论“本体”的三种含义及其现代混淆[J].哲学研究,2020(09):38-48+128.
- [8]Fernández-Izquierdo Alba,García-Castro Raúl. Conformance testing of ontologies through ontology requirements[J]. Engineering Applications of Artificial Intelligence,2021,97.
- [9]Arafeh Mohamad,Ceravolo Paolo,Mourad Azzam,Damiani Ernesto,Bellini Emanuele. Ontology based recommender system using social network data[J]. Future Generation Computer Systems,2021,115.