

《智能信息处理》课程考试

## 基于本体的教育资源检索研究

张锦磊

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

2021 年 12 月 19 日

# 基于本体的教育资源检索研究

张锦磊

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

**摘要:** 随着互联网的飞速发展, 越来越多的教育资源供我们选择, 资源的数量非常巨大。信息检索 (Information Retrieval) 是用户进行信息查询和获取的主要方式, 是查找信息的方法和手段。如何在如此多的教育资源中快速有效地检索到自己需要的资源是当前在线教育建设的难点。本文使用本体技术与信息检索相结合, 提出一种基于本体的教育资源检索系统。

**关键词:** 本体; 信息检索; 教育资源;

## Research on Educational Resource Retrieval System Based on Ontology

Zhang Jinlei

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026, Liaoning)

**Abstract:** With the rapid development of the Internet, more and more educational resources are available for us to choose from, and the number of resources is huge. Information Retrieval (Information Retrieval) is the main way for users to query and obtain information, as well as a method and means to find information. How to quickly and effectively retrieve the resources one needs from so many educational resources is the difficulty of the current online education construction. This paper uses the combination of ontology technology and information retrieval to propose an ontology-based educational resource retrieval system.

**Keywords:** ontology; information retrieval; educational resources;

## 1 引言

互联网技术随着时代与科技的发展已经日益成熟, 并且逐渐的融入了我们的日常生活, 成为了重要工具。互联网的教育资源非常丰富, 网络上的信息掺杂在一起, 在使用时需要甄选出具有应用价值的信息, 为满足这一需求, 各大编程工程师研发设计出了多种信息检索系统。如何在海量的互联网教育资源中精准有效的检索到所需要资源成为在线教育建设的难点。信息检索是信息按一定的方式进行加工、整理、组织并存储起来, 再根据信息用户特定的需要将相关信息准确的查找出来的过程。针对传统检索难以精准的查询到所需教育资源的问题, 本文基于语义,

提出一种智能教育资源检索系统, 帮助用户精准有效的检索到所需的教育资源。<sup>[1]</sup>

## 2 本体

### 2.1 概念

“本体(ontology)”一词起源于哲学领域的“形而上学”, 来自于亚里士多德对事物存在本质的研究, 在哲学上被定义为存在论, 即对客观存在事物做出的系统性说明。Alexander 将本体研究分成了两个方面, 一方面是研究事物存在的本质, 另一方面是对客观对象理论定义的研究。目前, 被学术界广泛认可的本体定义还是 Studer 给出的定义: “本体是对共享的概念模型做出的明确的、形式化说明”。该定义指出了本体的四大

特征：概念模型、明确化、形式化以及共享性<sup>[2]</sup>

在计算机科学与信息科学领域，理论上，本体是指一种“形式化的，对于共享概念体系的明确而又详细的说明”。本体提供的是一种共享词表，也就是特定领域之中那些存在着的对象类型或概念及其属性和相互关系；或者说，本体就是一种特殊类型的术语集，具有结构化的特点，且更加适合于在计算机系统之中使用；或者说，本体实际上就是对特定领域之中某套概念及其相互之间关系的形式化表达（formal representation）。本体是人们以自己兴趣领域的知识为素材，运用信息科学的本体论原理而编写出来的作品。本体一般可以用来针对该领域的属性进行推理，亦可用于定义该领域（也就是对该领域进行建模）。此外，有时人们也会将“本体”称为“本体论”。

作为一种关于现实世界或其中某个组成部分的知识表达形式，本体的应用领域包括（但不限于）：人工智能、语义网、软件工程、生物医学信息学、图书馆学以及信息架构。

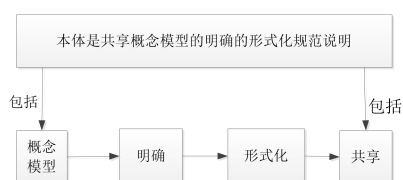


图1 本体概念描述

## 2.2 构成要素

常见的本体构成要素包括：个体（实例）：基础的或者说“底层的”对象。类：集合（sets）、概念、对象类型或者说事物的种类。属性：对象（和类）所可能具有的属性、特征、特性、特点和参数。关系：类与个体之间的彼此关联所可能具有的方式。函数术语：在声明语句当中，可用来代替具体术语的特定关系所构成的复杂结构。约束（限制）：采取形式化方式所声明的，关于接受某项断言作为输入而必须成立的情况的描

述。规则：用于描述可以依据特定形式的某项断言所能得出的逻辑推论的，if-then（前因—后果）式语句形式的声明。公理：采取特定逻辑形式的断言（包括规则在内）所共同构成的就是其本体在相应应用领域当中所描述的整个理论。这种定义有别于产生式语法和形式逻辑当中所说的“公理”。在这些学科当中，公理之中仅仅包括那些被断言为先验知识的声明。就这里的用法而言，“公理”之中还包括依据公理型声明所推导得出的理论。事件(哲学)：属性或关系的变化。<sup>[3]</sup>

## 2.3 描述语言

本体一般都是采用本体语言来编制的。本体语言，又称为“本体论语言”，是一种用于编制本体的形式化语言。概括地讲，一些主流的本体建模语言原则上应该具备以下特点：具有清晰准确的语法规则；语言交互性较好，兼具知识表达的便捷性与充分性；相关建模语言能够支持不同领域隐含知识的挖掘与推理。存在着许许多多此类本体语言，既包括专有的，也包括基于标准的。下面，对一些比较主流的本体建模语言展开介绍：

（1）RDF：RDF 是万维网联盟提出的一种用于描述领域知识的资源框架。它提供了节点与弧的描述结构，用以完成语义关系的表达。相比于 XML 语言，RDF 将领域中复杂的知识表示拆解为简单的三元组形式，能够以清晰明确的形式完成语义建模。RDF 与 XML 类似，在语义建模方面都有一定的局限性。它们没有对所描述的资源关系中的定义域和值域形成一定的约束关系，因而在复杂知识结构的表达上存在缺陷。为了弥补这方面的不足，RDF-S 应运而生。它对描述的知识结构进行了限制，并提供了相应的词汇表，有效地解决了复杂知识结构中的语义冲突等问题。

(2) OIL: OIL 是欧洲一些学者在 1999 年针对要完成的 IST 项目提出的一套本体建模语言体系。OIL 语言集兼容不同的知识结构的语义表达, 为不同领域的本体提供了相适应的建模语言。它的建模子语言集如下图 2 所示。由于其出色的性能与较好的知识表达能力, 得到了广大学者的高度关注与参与。

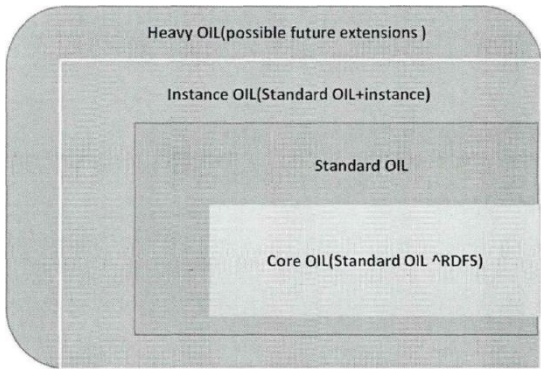


图 2 OIL 本体建模语言集

(3) OWL: OWL (是由美国的一些学者在 2000 年以后结合 DAML+OIL 描述语言的优势提出的一套新的本体构建语言。目前, OWL 语言处在本体建模技术栈的顶端, 是目前最为主流的建模语言之一。它根据知识工程的不同需求, 提供了三种不同的子语言用于知识的建模与描述。三种子语言口分别为 OWL Lite, OWL Full, 和 OWL DL。其区别与联系如下表 1 所示。本文围绕中医领域展开建模, OWL 语言自然成为了首选。

表 1 OWL 子语言介绍

子语言	描述	例子
OWL Lite	表达能力较弱, 适用于分类层次简单的本体模型	支持基数, 允许基数 0 或 1
OWL DL	约束条件高, 推理表达能力强	特定领域的本体构建中, 一个类被约束只能有一个父类
OWL Full	知识表达能力强。但取消了特定的约束条件, 不能保证对隐含知识可以判定推理	特定领域的本体构建中, 一个实例可以表示为不同类集合中的个体

(4) DAML: DAML 由美国学者在 2000 年提出的一套语义建模语言。它针对一些主流本体建模语言如 RDF 在知识表达方面的缺陷, 进行重新设计与改进。针对不同领域中复杂的知识结构, 它提供丰富的上下位关系, 并对知识网络中定义域与值域的描述属性进行增加限制, 完成复杂知识的组织与表达。DAML 语言和 OIL 语言在当时相比于传统的描述框架, 它们具有更为灵活的表达形式和推理能力。很快一套兼具各自优势的 DAML+OIL 的资源描述语言被提出, 成为了当时在知识建模方向最为主流的表达载体。<sup>[4]</sup>

2.4 本体构建

目前, 领域本体的构建还没有规范性的技术标准, 不同的工程领域对本体的应用有着不同的考虑, 因此知识本体的构建过程也就不尽相同。大致可分为四个阶段: 第一阶段是概念的抽取; 第二阶段为明确概念层次关系; 第三阶段是确定概念属性; 第四阶段为本体实例化。常用的本体构建方法有七种: TOVE 法、METHONTOLOGY 法、骨架法、KACTUS 法、SENSUS 法、IDEF5 法以及七步法。

骨架法: 骨架法是英国爱丁堡大学的学者提出的一套本体构建方法。其构建流程如图 3 所示。七步法: 七步法最早是由斯坦福大学的学者在进行医学本体领域构建时提出的一套方法体系, 其构建流程如图 4 所示。目前, 该方法被广泛地应用于各个领域的本体建模工程中。

TOVE 法: 主要用于构建企业本体、工程本体、服务本体以及计划本体,使用一阶逻辑谓词进行集成, 由人工构建, 没有完整生命周期。METHONTOLOGY 法: 主要用于构建化学领域知识本体, 由人工构建, 有完整的生命周期, 所构建的知识本体可用于多个领域。KACTUS 法: 主要用于解决复杂技术系统的知识复用, 主要用于电子网络等领域, 没有构建周期, 支持

EXPRESS 和 Ontolingua 语言, 使用 CML 语言对本体进行描述。SENSUS 法: 主要用于构建自然语言处理的 SENSUS 本体, 目前, 已被应用于构建电子科学、军事等领域的本体, 没有构建周期。IDEF5 法: 主要用于构建企业本体, 由人工构建, 没有构建周期, 利用图表和细节语言获取事实的概念、属性以及关系。

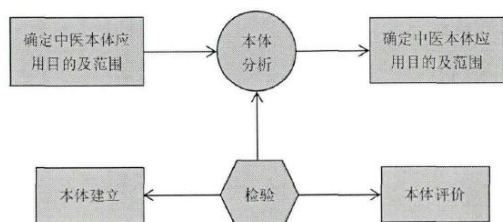


图3 骨架法构建流程

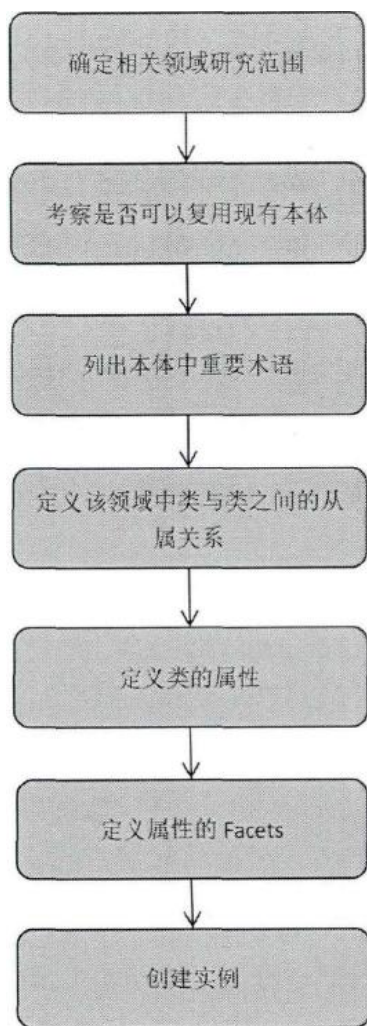


图4 七步法本体构建流程

### 3 基于本体的教育资源检索系统

基于本体的教育资源检索系统综合应用本体理论, 语义网络技术和信息检索等相关技术。基于本体的检索是通过的构建描述某个事物及其关系的具体的本体, 并且将其应用到相应的信息资源检索之中。具体在教育资源检索系统中, 首先需要有对教育资源及其相关关系进行定义和描述的本体, 然后基于教育资源本体来进行语义的推理和解析, 实现精确检索。

基于本体的教育资源检索系统由用户交换、检索引擎模块、教育资源本体模块、元数据库模块、网络采集模块和 RDF 实例模块六个部分组成。其中用户交互模块提供友好的用户查询检索页面; 检索引擎模块负责根据教育资源本体信息解析用户的查询请求, 并向元数据库请求用户所需数据, 并返回给用户; 教育资源本体模块定义教育资源本体内容, 在查询检索过程中负责语义分析和推理; 网络采集模块负责采集网络信息; RDF 实例模块负责对所采集的信息基于教育资源本体信息进行标注处理, 生产相应的本体实例, 并且形成对应的 RDF 实例文件; 元数据模块负责存储生成的 RDF 实例文件。基于本体的教育资源检索系统的系统结构如图 5 所示。<sup>[5]</sup>

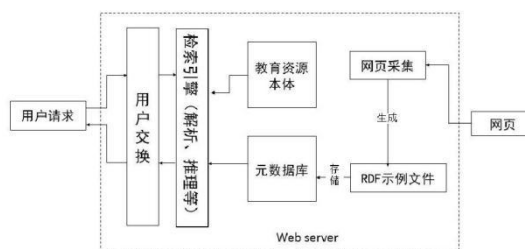


图5 基于本体的教育资源检索系统结构图

#### 3.1 本体建立

基于本体的语义检索模型设计的首要步骤就是本体的建立环节, 同时, 该环节也是系统设计的基础环节。基于本体的教育资源检索系统首先需要创建教育资源本体, 包含设计教育资源本体和将其应用到检索系统中。本体是一种形式化表示“知识”的工具, 目前已经广泛应用在语义网

络、信息检索、人工智能等领域。本系统设计的本体是应用在教育资源智能检索中,虽然有许多理论研究者与实践工作者将本体应用到教育领域,但是当前在网络教育资源领域还没有一个统一的标准和方法,因此需要根据具体的应用来进行设计。<sup>[6]</sup>本文采用本体编辑工具 Protégé 进行本体模型构建,包括定义本体的类、类属性、类层次和实例,设计教育资源本体的主要步骤包含如下几步:1) 确定范围,即确定教育资源本体的领域知识所包含的范围;2) 描述概念,描述教育资源本体的中的重要概念;3) 确定类,定义教育资源本体中的类和类的关系;4) 定义属性,定义本体类的属性。5) 确定属性约束;6) 建立实例。<sup>[5]</sup>

### 3.2 教育资源的获取

在模型的实际设计环节中,设计者可以充分利用网络爬虫将网上有利的教育资源下载下来,例如教育资源中的课件、网址资源以及名词术语等参考资源,并最终通过 Lucene 对预处理之后的网络资源进行索引库的存储。

### 3.3 语义标注

根据检索条件,基于教育资源本体,采用 jena 推理机进行语义推理,最终得到检索结果。在创建了教育资源领域信息知识本体构和语义标注后,使用 Jena 推理机将 XML 文件转化为可供检索程序应用的格式。<sup>[7]</sup>Jena 是一款开源的,由惠普公司开发的基于 Java 的语义网工具包,是用来创建语义 Web 系统的框架,它提供基于本体的语义检索相关工具 API。<sup>[8]</sup>

### 3.4 信息检索

根据检索条件,基于教育资源本体,采用 jena 推理机进行语义推理,最终得到检索结果。在创建了教育资源领域信息知识本体构和语义标注后,使用 Jena 推理机将 XML 文件转化为可供检索程序应用的格式。<sup>[9]</sup>Jena 是一款开源的,由惠普公司开发的基于 Java 的语义网工具包,是

用来创建语义 Web 系统的框架,它提供基于本体的语义检索相关工具 API。

## 4 结语

在教育资源越来越多的今天,如果可以准确快速的找到自己想要的內容会给我们节省很多的时间。本文设计的基于本体的教育资源的信息检索模型,与传统教育资源检索相比,基于本体的网络教育资源检索通过教育资源的本体信息在本体层来检索教育资源,利用合理完善的教育资源本体描述,表达了教育资源概念的语义含义和相互之间的内在联系,帮助用户更精准的检索到所需要的教育资源信息,更好地满足用户需求。

### 参考文献

- [1] 黄以宝.基于本体的教育资源语义检索系统的实现探讨[J].信息与电脑(理论版),2018(18):182-184.
- [2] 杨超. 基于本体的微派建筑知识图谱构建研究[D].安徽建筑大学,2021.DOI:10.27784/d.cnki.gahjz.2021.000155.
- [3] 张丽伟. 基于本体的学习资源个性化推荐系统研究[D].东北石油大学,2019.DOI:10.26995/d.cnki.gdqsc.2019.000242.
- [4] 曹睿杰. 面向中医领域的本体构建与知识发现应用研究[D].北京邮电大学,2021.DOI:10.26969/d.cnki.gbydu.2021.001921.
- [5] 刘媛媛.基于本体的教育资源检索系统研究[J].电脑知识与技术,2017,13(35):1-2.DOI:10.14004/j.cnki.ckt.2017.4038.
- [6] 林丽姝,林珍,刘露思.基于本体与粗糙集的教育资源检索模型的关键技术研究[J].实验技术与管理,2012,29(10):123-125.DOI:10.16791/j.cnki.sjg.2012.10.034.
- [7] 栾艳,丁二玉,骆斌.基于 Ontology 的语义检索技术[J].计算机工程与应用,2005(28):156-159.
- [8] 刘志亮,张金隆.基于教育云的教育信息资源整合框架研究[J].情报杂志,2013,32(10):164-167.
- [9] 万静,王文聪,易军凯.一种基于本体的知识库语义扩展搜索方法[J].计算机工程,2012,38(06):19-21+24.