

基于信息抽取的课程知识链接方法

文必龙,马 强,李 菲

(东北石油大学 计算机与信息技术学院,黑龙江 大庆 163318)

摘 要:课程教材资源存在分散无序、知识点聚合性不足、经常变化的问题,同时传统的依靠人为的教材资源组织方法,存在耗时费力、组织结果不易更新等问题。为了实现对课程教材资源中知识点的有效组织和利用,提出了课程知识链接描述模型,对课程资源中分散的碎片化知识进行描述,然后采用信息抽取方法,对课程教材资源中的知识点及知识点之间的关联关系进行抽取,将抽取的知识点进行标识,计算教材资源中文档与知识点间的关联,最终建立课程教材知识链接体系。以教育心理学课程为例,建立知识链接原型系统,抽取了领域内概念435个,并对各文本内容进行了标注,实现了该方法的应用。方法在不改变现有的资源组织架构的前提下,实现知识点与知识点、知识点与文档的关联,具有很强的适应性和可用性。

关键词:教材资源;信息抽取;知识链接;知识关联;知识文档关联

中图分类号:TP391.1

文献标识码:A

文章编号:1673-629X(2020)04-0146-06

doi:10.3969/j.issn.1673-629X.2020.04.028

Method of Linking Course Knowledge Based on Information Extraction

WEN Bi-long, MA Qiang, LI Fei

(School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

Abstract: There are some problems in the resources of curriculum teaching materials, such as scattered disorder, insufficient aggregation of knowledge points and frequent changes. At the same time, the traditional method of relying on artificial teaching material resource is time-consuming and laborious with the organization result not easy to update. In order to realize the effective organization and utilization of knowledge points in curriculum textbook resources, a course knowledge's link description model is proposed, and the scattered fragmentation knowledge in curriculum resources is described. Then the relationship between knowledge points and knowledge points in the course textbook resources are extracted by the information extraction method. The extracted knowledge points are identified, the relationship between Chinese file and knowledge points of textbook resources is calculated, and finally the knowledge's link system of course textbook is established. Taking the course of educational psychology as an example, the prototype system of knowledge link is established, 435 concepts in the field are extracted, and the contents of each text are labeled, realizing the application of this method. Without changing the existing resource organization architecture, this method realizes the relationship between knowledge points and knowledge points, knowledge points and documents, which has strong adaptability and usability.

Key words: teaching materials; information extraction; knowledge linking; knowledge correlation; knowledge document links

0 引言

随着教育信息化建设与教学过程的深度融合,以及云课堂、网络课程、移动学习^[1]等教学形式在教育领域中的广泛深入应用,产生并积累了大量的课程教材资源。这些资源中蕴含着丰富的领域性知识。然而这些课程教材资源中的知识存在分散无序、聚合性不足的问题,难以满足使用者对教材资源有效利用的迫切需求,如教师或学生在学习某个知识点时需要查找相

关的知识点,或者需要搜寻其他教材对同一知识点的讲解。因此,对课程教材资源中的知识进行描述与组织,建立知识点之间的链接、知识点和教材资源间的链接显得越来越重要。

知识链接是根据知识体之间的关联关系将它们联系起来的过程^[2]。传统的教育资源管理、教材分析方法、教育统计学等工作对知识链接的研究虽有一定的成就,但研究基本局限在文献单元层面或资源整理层

收稿日期:2019-05-11

修回日期:2019-09-16

网络出版时间:2019-12-05

基金项目:国家自然科学基金面上项目(41574117);国家重大专项(2016ZX05033-005-004)

作者简介:文必龙(1967-),男,博士,教授,研究方向为大数据、软件工程、信息集成;马 强(1993-),男,硕士研究生,研究方向为知识工程。

网络出版地址:<http://kns.cnki.net/kcms/detail/61.1450.TP.20191205.1146.074.html>

面,并且需要投入大量人力进行整理,难以应对课程教材资源不断增长的挑战。教材资源多数以半结构化和非结构的文本形式出现,需要相应的技术进行处理。当前,信息抽取技术在信息组织、文本非结构化到结构化转换等文本处理任务中取得了显著成功,它从给定的文档集合中自动识别出预先设定的实体、关系和事件等类型的信息^[3],并将这些信息进行结构化存储和管理。因此可借鉴其方法用于课程知识链接体系的建立当中。

文中提出基于信息抽取的课程教材资源知识链接方法,使其不限于数据体量,以实现大量课程教材资源进行处理。相比于传统方法,对提高课程知识点间的关联分析与教材资源的高效利用具有现实意义。

1 相关工作

构建课程知识链接体系源于对数字化课程教材资源进行整序、共享与利用的客观需求。文献[4]认为知识链接是指通过知识关联将具有相关关系的知识单元按照一定的需要有序地联系在一起,形成序列化或结构化的知识集合,继而构成知识网络的一种行为。文献[5]认为知识链接提供了一个知识环境,可从一个信息点切换到相关信息点,实现特定领域多种知识资源的有机关联。在构建知识链接体系的实践中,研究者采用不同的方法进行链接的建立。文献[6-7]提出基于引文的关联检索,整合不同学科的知识源,实现科学文献间的知识链接,在语义 Web 环境中构建知识网络。文献[8]实现基于用户行为的知识链接向用户提供知识服务。文献[9-10]提出基于语义的链接,使数字图书馆资源的语义能被机器理解。其中,基于知识元的知识组织链接在实际应用中取得了很大成功,它基于资源间的内在关联,构建概念关系、层次信息等的知识结构网络,更加强调知识结构的动态发现与多维聚合。典型的基于知识元的应用有学习元平台、基于标签的知识资源聚合、中国知网概念知识元库等。从相关研究与实践可以看出,知识链接体系有两个基本要素:知识单元及知识单元之间的关联关系。

在课程教材资源的知识链接体系构建中,对知识单元的认识和知识单元结构的描述是链接构建的基础和信息抽取的依据。一般认为“知识点”是教材中教学内容的最小单元。文献[11]认为应以知识点下辖的知识元为组成教学知识点的最小、独立的知识单位。情报学界的研究者认识到应该将信息标引深入到文献中的知识元层次,并提出知识元这一概念。温有奎教授把知识元当作知识结构的基本元素。在知识元结构表示形式方面,文献[12]提出了知识元的概念、关系、问题三元组表示。文献[13]提出了知识元的名称、属

性、属性描述和信息接口四元组表示。对知识结构的表示是建立链接构建的基础工作。

课程教材资源的信息抽取是以课程教材文档集合为输入,识别并抽取课程教材中的专有名词、命名实体及关系、事件等指定的信息,将抽取的信息进行结构化存储。文献[14]利用信息抽取技术,通过术语识别和标准化等工作建立了结构化的药物方剂知识库。文献[15]利用信息抽取技术对 Web 资源进行抽取,以建立移动学习资源库。在课程知识的链接工作中,领域概念(以下所述“概念”均指课程领域专有名词及专有命名实体)是构成知识单元的最基本要素,领域概念和领域概念间的关联关系是信息抽取的重点。

2 方法

2.1 总体流程

针对课程教材资源中知识点分散无序、聚合性不足的问题,文中提出课程知识链接体系描述模型,基于该模型设计了课程知识链接体系建立的流程,主要包括:课程教材采集、教材信息处理、概念及概念关联抽取、概念与文本关联计算,如图1所示。

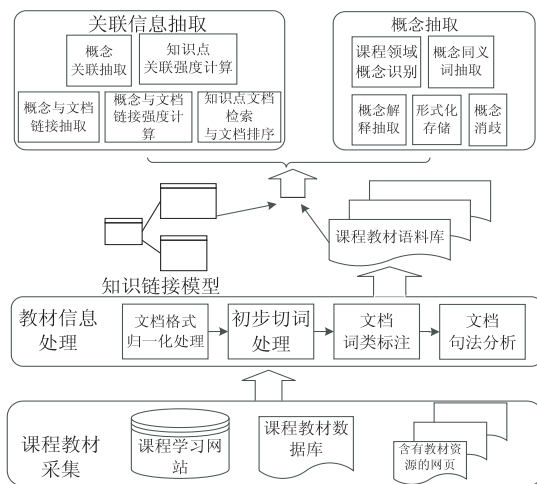


图1 课程知识链接建立流程

课程教材采集是针对每类数据源开发一种爬虫器,从课程学习网站、课程教材数据库、含有教材资源的网页等数据源中爬取大量文档。教材信息处理将爬取得到的各种格式文档转换为统一的文本格式,并对文档内容进行分词、词类标注、句法分析等自然语言处理,构建课程语料库,为概念及概念关联抽取、概念与文本关联抽取做准备。概念及概念关联抽取是对文档中的领域专有名词、领域命名实体进行识别并发现概念间的关联的过程。基于课程知识链接模型,采用规则匹配与统计学习相结合的方式,抽取课程概念、挖掘概念间的关联关系并存储。最后抽取概念和文档间的关联信息,包括概念与文档链接信息的抽取和概念与文档链接强度的计算两部分。概念与文档链接信息的

抽取主要是实现概念与文档的检索排序,概念与文档链接强度的计算是衡量两者的隐含关联程度。上述课程知识链接体系构建的核心工作是概念及概念间关系的抽取、概念与教材文档间关联信息的抽取。

2.2 课程知识链接描述模型

课程知识链接描述模型是对课程教材中的知识点、知识点的特征及知识点间的联系进行抽象的工具,也是课程知识链接体系的基础与核心,该模型定义如下:

定义 1:课程知识链接体系是课程知识点及课程知识点间关联关系的集合,可以描述为一个 2 元组:

$$\text{Knowledge_Linking} = \{ K, R \}$$

其中, K 为课程知识点 (knowledge) 的集合; $R = \{ r_1, r_2, \dots, r_n \}$ 为课程知识关联关系 (relevance) 集合。

定义 2:课程知识点 (Knowledge) 是教材资源中信息传递的基本单元,表示为:

$$\text{Knowledge} = \{ \text{Concept}, \text{Alias}, \text{Type}, \text{Definition} \}$$

其中, Concept 为该知识点的概念,用词表示; Alias 为概念的同义词列表; Type 为概念类型; Definition 为课程知识点概念的基本解释。

定义 3:知识点之间的关联关系 (Relevance) 是以某一中介为纽带建立的联系,表示为一个 4 元组:

$$R = (K_1, K_2, \text{Type}, \text{Value})$$

其中, K_1, K_2 表示知识点; Type 表示知识点 K_1 和知识点 K_2 间关系的类型,例如语义相关类型、共现相关度类型等; Value 表示某关系类型下的具体关系取值。

定义 4:课程教材中的章节 (Section) 表示为:

$$\text{Section} = \{ \text{Name}, \text{Book}, \text{Serial}, \text{Content} \}$$

其中, Name 表示该章节在教材中的名称; Book 表示该章节所属的教材; Serial 表示该章节在教材中的编号; Content 表示该章节中的具体内容。

2.3 课程知识点抽取

2.3.1 知识点概念识别

课程教材资源中包含的课程知识点概念以专有名词、专有命名实体的形式出现。利用信息抽取中的固定短语抽取、命名实体识别技术对教材领域概念进行抽取。抽取方法结合了领域知识与半监督学习方法。

文中从以下途径获取领域知识:(1) 课程领域已有的知识库,包括教育大词典和教育主题词表;(2) 特殊语料库,包括维基百科词条、中国大百科数据库。从中提取课程中概念术语及其关系,构建课程领域知识库。再利用得到的领域知识生成领域空间辞典和领域本体,最后使用辞典和本体对课程教材中的词汇序列标注为领域概念。这部分功能集成自前期工作^[16-17]。

由于缺乏标注的领域课程语料库,因此文中采用半监督学习方法进行处理,使用开源工具 CRF++ 进行潜在的领域领域的实体抽取,并确定实体类型。选取的特征有上下文范围内的词、词性、词频和句法特征等。采用互信息与左右信息熵对未登录到课程领域词典中的专有名词进行识别与标注。

2.3.2 概念同义词抽取

同义词抽取的目标是寻找表示同一概念的术语。文中总结了同义词表达的基本模式,基本模式使用的特征词见表 1。

表 1 同义词抽取所用特征词

序号	第一类	第二类
规则	简称、简称为、全称、被称为、乃称为、称作、亦称、亦可称、又称、故称、古称、即称、俗称、人称、世称、初称、泛称、特称、原称、代称、旧称、异	即“[...]”、亦即、又名、亦名、改名为、现名、原名、改现名、初名、改此名、定名、命名为、
特征词	称、改称、后称、别称、或称、或者称、便称之为、略称、统称、总称、合称、均称、通称、有[...]之称、史称、时称、始称	故名、得名、别名、又指、又叫、又作、一译、旧译、也译、又译

然后利用模版匹配法,根据预先设定的同义词抽取模版,将教材信息处理阶段得到的课程语料与同义词模版中的规则依次进行比较,如果匹配成功,则提取同义词,加入到同义词词典中。最后再根据输入到系统中的概念,从构建好的同义词词典中返回与其含义相同的同义词。

2.3.3 概念解释抽取

概念的解释是指对此概念描述对象的正式或非正式的定义,如“所谓学习动机是指直接推动学生进行学习活动的内在心理因素或内在动力”就是对学习动机的定义。分析了课程教材文本中解释文本的特点,总结了基于模式的抽取规则和基于语法的抽取规

则,综合利用两类规则对概念的解释进行抽取。部分解释抽取规则见表 2。

2.3.4 知识点存储

完成上述抽取后,对知识点进行形式化存储。存储形式依据课程知识点模型,并添加唯一的知识点识别号。

2.3.5 概念消歧

消歧的目的就是明确概念的语义。由于不同课程教材资源中存在语义不明确问题,如:布鲁纳提出的“发现学习”阐述的是学生主动获得新知识,而奥苏贝尔的“发现学习”阐述的是学生发现新旧知识之间的关系。因此在知识点链接系统中集成消歧功能。

表2 概念解释抽取规则

类型	模式规则	语法规则
内容	1. (所谓 认为 null)(Tem)(是 是指 就是 定义是 指的是)…(的)	1. (np n(include Term))(include 是)(adj n np sen)(的)(n)
	2. (所谓 认为 null)(Term)…(定义 是)…	2. (np n(include Term))(include 是)(一种 一类 一个)(n)(,)(sen)
	3. (凡是 只要 如果 null)…(的)…(即为 称为 是)(Term)	3. (adj n np sen)(的)(n)(即为 称为 include 是)(np n(include Term))

利用 2.3.1 中得到的领域概念词典和本体进行消歧,消歧的方法如下^[18-20]:针对有歧义的课程知识概念 w ,首先取得其在不同语料资源中的每种可能解释,利用本体计算得到语义子图 $\text{Graph} = \{(\text{node}_1, \text{node}_2, \dots, \text{node}_n), (\text{relation}_1, \text{relation}_2, \dots, \text{relation}_m)\}$;然后依次计算 w 的每个语义子图与该概念目前所在的上下文的语义子图 Context-G 间的相似度;最后取相似度最大的图对应的概念作为 W 的确定的语义。

2.4 课程知识点关联度抽取

课程概念间的关联强度,通过综合计算概念在课程知识库中的语义信息和概念在课程资源语料集中的共现信息得到。

2.4.1 概念间语义关联度计算

词嵌入方法在计算文本语义信息上非常有效。利用课程教材语料训练词向量,将词向量用于领域概念语义相关性的计算上。

设知识点的概念 x_i 与知识点的概念 x_j 的 k 维向量为: $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ 、 $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jk})$,则词 \mathbf{x}_i 和词 \mathbf{x}_j 的语义相关度为:

$$\text{semStrength}(x_i, x_j) = \frac{\left| \sum_{m=1}^k x_{im} * x_{jm} \right|}{\sqrt{\left(\sum_{m=1}^k x_{im}^2 \right) \left(\sum_{m=1}^k x_{jm}^2 \right)}}$$

2.4.2 概念间共现关联度计算

在知识链接系统中,以概念抽取阶段得到的领域概念词表为词语基表,以课程教材语料处理后的自然段为窗口单位,统计基表词语共现频次,得到共现矩阵 CoF 。

概念本身的关联度计算如下:

$$R_i = \sum_{j=1}^n \text{CoF}_{ij} (i \neq j)$$

其中, i 表示概念共现矩阵中行序或列序为 i 的词; j 表示概念共现矩阵中行序或列序为 j 的词; CoF_{ij} 表示概念 i 和概念 j 的共现频次。

概念间的关联度计算如下:

$$\text{coSimilarity}(x_i, x_j) = \frac{\text{CoF}_{ij}}{\text{MaxF}}$$

其中, MaxF 为共现矩阵中最大的共现词频。
相关事物通常会同时出现,同时出现的事物往往

也具有一定相关性^[21]。两个或更多词在同一单位中同时出现称为词共现,共现的词之间必定具有一定的关系。词共现关系在揭示相关性知识,挖掘隐性关联知识方面都有重要意义。

2.4.3 概念关联强度计算

综合课程概念语义关联度和课程概念共现关联度可得知识点链接强度:

$$\text{LinkStrength}(x_i, x_j) = \alpha * \text{coSimilarity}(x_i, x_j) + \beta * \text{semStrength}(x_i, x_j)$$

其中, α 和 β 为可调节权重, $\alpha + \beta = 1$ 。

将知识点链接强度的计算结果与领域知识库中抽取的概念关联组合,可以用于生成知识点链接。

2.5 知识点与教材文档间关联抽取

2.5.1 概念与文档链接信息抽取

以包含课程知识点概念和概念关联的知识库为基础,依据课程知识链接描述模型,在教材语料中抽取描述知识点的信息,建立课程知识点概念与文档间链接。

首先,从特定教材文档中识别已有概念,依托概念集合对文档进行扫描和标识,将文档中概念词语链接到其文档。其次,建立从概念到文档的关联,即:有哪些文档包含该知识点。该功能利用 Lucene 检索工具实现。

2.5.2 概念与文档链接强度计算

将教材内容以章节为单位切分然后存储,使用 LDA 模型计算知识点和该节的链接强度。LDA 是一种统计模型,可以计算概念和文档间的相关性。首先将章节内容进行分词,然后计算分词后词语在主题上的概率分布,计算方法如下:

$$P(w_i | T_k) = \frac{C_{ik} + \beta}{\sum_{i=1}^N C_{ik} + N * \beta} = \psi_i^{t=k}$$

其中, C_{ik} 表示词 i 被赋予主题 k 的次数。

之后计算主题在文档上的概率分布,计算方法如下:

$$P(T_k | D_m) = \frac{C_{km} + \alpha}{\sum_{k=1}^K C_{km} + K * \alpha} = \theta_{t=k}^m$$

其中, C_{km} 表示文档 m 中词被赋予主题 k 的次数。
最后,计算词语在文档上的主题重要性,即概念与

文档间的链接强度,计算方法如下:

$$P(w_i | D_m) = \sum_{k=1}^k \psi_i^{t=k} * \theta_{t=k}^m$$

以上公式中, α 和 β 表示超参数。用于平滑非包含的单词和主题。

3 原型实现

基于以上所述流程,对教育心理学的教材进行实验,构建了教育心理学课程知识点链接系统原型。

3.1 课程知识抽取与计算

(1)教材文本采集。

教育心理学课程具有领域知识数量大、难理解的特点,原有课程的教学资源以文档为粒度组织,难以被学习者充分利用。教材信息采集阶段共搜集不同的课程教材 25 种。

(2)领域知识选择。

为了使用领域辞典和本体识别领域概念,采用了中国大百科教育类、心理学类词条及顾明远主编的《教育大辞典》为领域数据来源构建领域本体。进入中国大百科,筛选“教育”、“教育心理学”类;采集词条 URL 内容,以 HTML 代码匹配方式提取词条信息框内容;提取信息框相关词条下列的词语转化为与该词条对应的细粒度概念关联,共得到词条 1 752 个。从《教育大辞典》中抽取教育类词条 34 103 个。

(3)文本预处理。

教材信息处理阶段,核心工作使用哈工大语言技术平台 LTP 进行,将搜集的领域词条添加到词典中,对教材文本进行分词及词类标注、命名实体识别。

(4)课程知识抽取与计算。

抽取工作按照 2.3、2.4、2.5 所陈述的方法进行。其中词向量的训练使用 word2vec 开源工具处理,使用中先利用分词工具对语料进行分词,分词时,加入获取到的领域概念词典,提高分词准确性。去除停用词及词频小于 10 的词语,减小训练的规模。训练时词语的维度设置为 200 维。

3.2 功能实现

课程知识链接原型系统采用 B/S 方式实现,Java 语言开发。数据库采用 Oracle11g。可视化部分集成 ECharts 组件。系统主要功能展示如下:

(1)知识点检索。

从数据库中检索抽取好的知识点信息,信息包括该知识点的同义词信息,介绍该知识点的教材有哪些,在这些教材中是如何解释该知识点的,等等,如图 2 所示。

(2)知识点关联词。

知识点关联词检索,用于检索与查询的知识点最

相关的知识点有哪些,每个知识点与查询的知识点的相关程度有多大。可帮助学习者全面掌握知识,也可判断其他知识点对查询知识点的重要性,如图 3 所示。

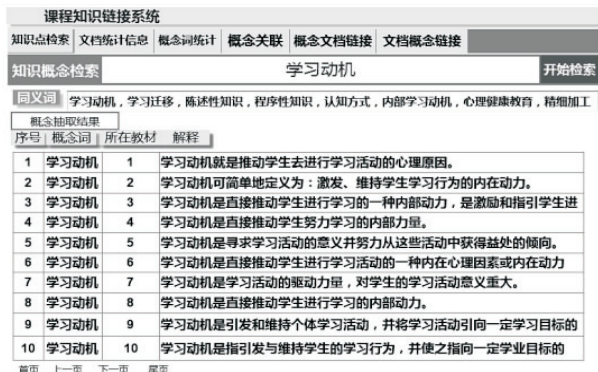


图 2 知识点检索模块

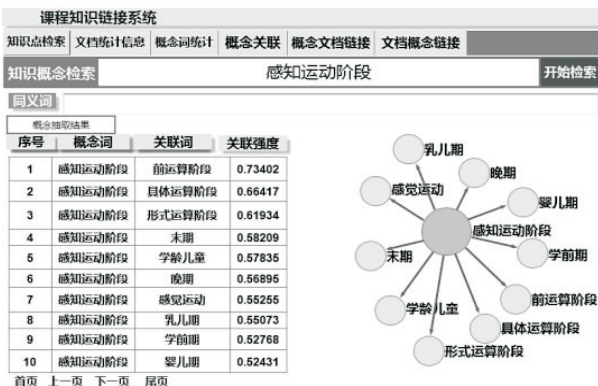


图 3 知识点关联查询模块

(3)知识点文档关联。

检索文档中主要介绍的内容是什么,文档中包含的主题,文档中涉及的主要知识点,如图 4 所示。

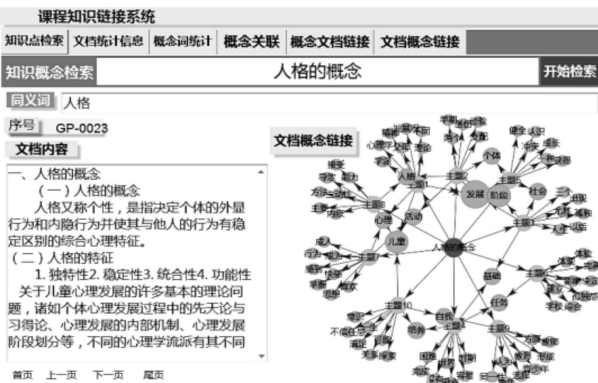


图 4 文档知识,点关联查询模块

此外,在使用过程中,可根据需要进一步增加功能模块,如基于知识链接网络发现核心知识、边缘知识,按重要性对知识点进行排序等。

4 结束语

教材数量快速增长,教材资源越来越丰富,实现课程知识和教材的连接、连通是资源得以高效利用的关键。通过构建知识点组织描述模型及采用信息抽取技

术,实现课程资源中的知识点的关联、知识点和教材文档的关联。学习者可通过知识点链接实现从文档到知识点、知识点到知识点、知识点到文档的连通,完成对相关知识内容的横向了解和纵向深入学习。方法不改变课程资源中现有的资源组织架构,具有很强的可用性和适应性。教材资源的链接构建是一个长期的应用实践过程,文中提出的从教材信息采集、处理、抽取、组织的方案能够实现教学资源的基于概念的关联,减少教材资源组织过程中人力、物力的大量投入,对资源的组织具有重要的参考意义。

同时,该工作还存在以下问题有待继续研究:(1)方法目前是在领域知识支撑的情况下进行的,可以实现细粒度的知识点抽取,而某些领域中领域知识较少,如何对缺少领域知识支持的教材资源进行处理,还需对方法进行完善;(2)只关注了知识点间以概念为主的链接,而知识单元可从更大粒度的表示形式去表示,如何解决这样的知识链接问题,还有待深入研究。

参考文献:

- [1] TING Y. Using mobile technologies to create interwoven learning interactions: an intuitive design and its evaluation [J]. *Computers & Education*, 2013, 60(1): 1-13.
- [2] 周晓英. 知识链接的发展阶段、发展动因和类型特征分析 [J]. *图书情报工作*, 2010, 54(12): 36-40.
- [3] 郭喜跃, 何婷婷. 信息抽取研究综述 [J]. *计算机科学*, 2015, 42(2): 14-17.
- [4] 曾建勋, 赵捷, 吴雯娜, 等. 基于引文的知识链接服务体系研究 [J]. *情报理论与实践*, 2009, 32(5): 1-4.
- [5] 贺德方, 曾建勋. 基于语义的馆藏资源深度聚合研究 [J]. *中国图书馆学报*, 2012, 38(4): 79-87.
- [6] HU Z, FANG S, LIANG T. Empirical study of constructing a knowledge organization system of patent documents using topic modeling [J]. *Scientometrics*, 2014, 100(3): 787-799.
- [7] CHEN C, WU Z, WU D, et al. Nearest access: rapid location for interlibrary loan resources using OpenURL [J]. *The Electronic Library*, 2011, 29(6): 792-802.
- [8] 廖秋荣. 基于用户行为的数字图书馆服务组织研究 [J]. *图书情报导刊*, 2016, 1(8): 87-90.
- [9] 刘健. 数字图书馆资源聚合与服务推荐研究 [D]. 长春: 吉林大学, 2017.
- [10] ALONSO I, CONTRERAS D. Evaluation of semantic similarity metrics applied to the automatic retrieval of medical documents: an UMLS approach [J]. *Expert Systems with Applications*, 2016, 44(3): 386-399.
- [11] 占小红, 王祖浩. 师范生课堂知识结构的特征研究——基于课堂知识结构与教材知识结构的比较 [J]. *全球教育展望*, 2014, 43(8): 83-91.
- [12] 高国伟, 王亚杰, 李永先. 知识元表示方法研究 [J]. *现代情报*, 2015, 35(3): 15-18.
- [13] 姜永常. 基于知识元语义链接的知识网络构建 [J]. *情报理论与实践*, 2011, 34(5): 50-53.
- [14] 朱彦, 朱玲, 王俊慧, 等. 基于信息抽取的历代方剂药物知识发现方法及应用 [J]. *中华中医药杂志*, 2015, 30(5): 1447-1451.
- [15] 王小根, 邓烈君, 王露露, 等. 基于知识元的移动学习资源组织模式研究 [J]. *电化教育研究*, 2017(1): 86-93.
- [16] 文必龙, 段炼, 汪志群, 等. 基于语料库和规则库的石油本体自动构建研究 [J]. *计算机技术与发展*, 2015, 25(9): 209-212.
- [17] 文必龙, 李云静. 基于油田领域本体的信息抽取技术研究 [J]. *计算机技术与发展*, 2015, 25(7): 226-229.
- [18] BUSCALDI D, ROSSO P. A conceptual density-based approach for the disambiguation of toponyms [J]. *International Journal of Geographical Information Science*, 2008, 22(3): 301-313.
- [19] SCHADD F C, ROOS N. Word-sense disambiguation for ontology mapping: concept disambiguation using virtual documents and information retrieval techniques [J]. *Journal on Data Semantics*, 2015, 4(3): 167-186.
- [20] BARATHI M, VALLI S. Ontology based query expansion using word sense disambiguation [J]. *International Journal of Computer Science & Information Security*, 2010, 7(2): 866-886.
- [21] 浦墨, 郑彦宁, 赵筱媛, 等. 基于词共现关系强度和关键词词频的叙词选词方法探究 [J]. *图书情报工作*, 2013, 57(15): 121-125.