

Drupal 平台下关联数据在MOOC教育资源共享中的研究*

许艳丹,张前进

(安徽国防科技职业学院,安徽 六安 237011)

【摘要】通过对MOOC教育资源的应用现状与关联数据在教育资源共享应用中的应用现状研究,提出使用关联数据解决MOOC课程资源碎片化、片段化的解决方案,并介绍了关联数据应用的实现过程、体系框架。开源CMS内容管理平台Drupal对关联数据的支持使预算较小的MOOC平台进行教育资源共享成为可能。最后介绍了利用Drupal实现创建与发布关联数据。

【关键词】关联数据;内容管理系统;Drupal;MOOC

【中图分类号】TP391;G434 **【文献标志码】**A **【文章编号】**1673-1891(2015)04-0046-04

0 引言

从2001年麻省理工学院实施的“开放式课件”(Open Courses Ware, OCW)项目到2002年联合国教科文组织在巴黎召开的“开放课件对于发展中国家高等教育发展的影响”专题论坛上提出的“开放教育资源(Open Educational Resources, OER)”概念,开放获取理念的运动已蓬勃发展了14年。网络上的开放教育资源(特别是开放课程)日渐丰富^[1]。2012年,大型开放式网络课程,即MOOC(Massive Open Online Courses)快速成长。MOOC的出现是开放式课程发展到新阶段的产物。MOOC是基于在线网络平台针对大规模不确定受众的学习平台,具有课程受众面广、参与自主性强、课程资源丰富、课程知识碎片化、课程服务个性化等特点。2012年以来,MOOC获得了飞速发展,比较有代表性的有Coursera、Udacity、edX、FutureLearn、Open2Study等MOOC平台提供商。这些平台提供的课程数量也是逐年成倍增加,如2014年4月Courdera提供600多门,edX提供近200门,到了2015年9月Coursera提供1300多门^[2],edX提供500多门^[3]。国内MOOC发展也是比较迅速,安徽省仅2014年省级资助立项MOOC达107门课程。这些数量巨大MOOC课程资源,一般是以松散性的知识点作为课程资源的组织方式。这样就造成了MOOC课程资源以一种碎片化的状态存在于各个MOOC平台中,难以被用户发现和重用。同时,碎片化的资源能够为用户提供更加多样化的学习,哈佛大学的Peter K.Bol教授提出“我们的MOOC平台反而要考虑如何提供更加碎片化和结构化的多样选择,比如以片段而不是课为单位^[4]”。因此如何将大量碎片化、片段化、结构化并

且具有相同主题的MOOC课程资源串联起来,方便用户进行课程检索,同时能根据用户的需求为其提供个性化推送服务是MOOC平台需要改进的方向。

关联数据描述了一种发布结构化数据的方法,该方法能够让数据相互关联,并通过语义查询让数据更有用。关联数据是建立在比如HTTP、RDF和URIs标准的Web技术上,但并不是使用这些技术为人类读者提供Web页面,而是扩展它们实现能够通过计算机自动阅读来分享知识的方法^[5]。通过关联数据的概念描述可以看出基于关联数据(Linked Data)的语义数据网络可以使各个MOOC平台中的课程资源有效关联契合在一起,形成一个巨大的数据Web,对于MOOC课程资源的发现、重用、分享、融合、互操作具有重要作用。本文通过对关联数据的研究,探讨基于Drupal的关联数据在MOOC教育资源Web应用开发中的应用。

1 关联数据在教育资源共享应用中的现状

目前最典型的数据关联应用项目是2007年启动的链接开放数据项目(Linking Open Data, LOD)。该项目目标是通过定义已存在并可公开使用数据集,根据关联数据原则将其转换为RDF并上网发布,进而改善数据网络(Web of Data)^[6]。由于该项目的开放性,参与者只要遵循关联数据发布原则进行发布数据集并将其与已有数据集相关联即可,因此项目进展迅速。图1显示了截至2014年8月LOD发布的数据集。LOD各个历史版本包含的数据集个数如表1所示。

收稿日期:2015-10-19

*基金项目:安徽省质量工程项目-网页设计与制作大规模在线开放课程(MOOC)示范项目(2014mooc088);安徽国防科技职业学院质量工程项目(gf2014xbjc10)。

作者简介:许艳丹(1983-),女,河南周口人,讲师,硕士,研究方向:数据挖掘、高职教育。

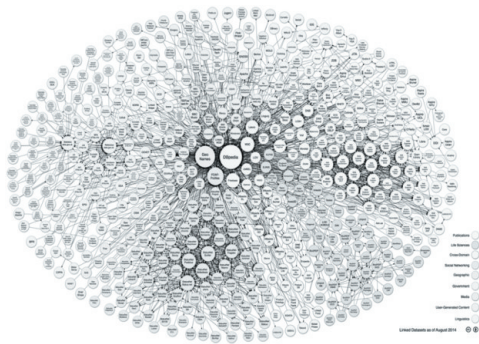


图 1 LOD 云图^[7]

表 1 LOD 部分历史版本^[7]

Version date	The number of data sets
2014-08-30	570
2011-09-19	295
2010-09-22	203
2009-07-14	95
2009-03-27	93
2007-05-01	12

根据 CKAN 的分析统计,截至 2011 年 9 月 LOD 云图中的关联数据集涵盖了新闻媒体、地理、政府、新闻出版、交叉领域、生命科学、用户内容生成等七大领域^[8]。尽管目前关联数据集涉及的领域较广,但是在教育资源领域应用的项目还不是很多。在实践应用上,国外发展较快。如吴鹏飞^[9]在国外关联数据教育应用项目研究与启示一文中介绍了 SemTech^[10] (Semantic Technologies for Learning and Teaching)、mEducator^[11]、LUCERO^[12]等三个国外关联数据在教育领域的典型应用平台;黄永文^[13]在关联数据应用的系统框架及构建关联数据应用的建议一文中提到的 Talis Aspire^[14]。目前国内关联数据教育资源领域的应用实践项目多集中在图书馆应用上,如上海市中心图书馆名录^[15]。

2 Drupal 平台下构建基于关联数据的 MOOC 教育资源共享平台

2.1 关联数据应用的体系框架

关联数据应用框架由关联数据发布层 (Publication Layer)、数据网络层 (Web of Data)、数据访问,集成和存储层 (Data Access, Integration and Storage Layer)、应用层 (Application Layer) 组成。关联数据应用框架如图 2 所示:

关联数据应用实现的过程:

第一步:通过关联数据发布层依据关联数据规则将数据通过标准的 HTTP 协议发布到数据网络层。关联数据的数据源可以是传统的结构化数据、基于 CMS 的节点属性描述的数据 (如 Drupal),也可

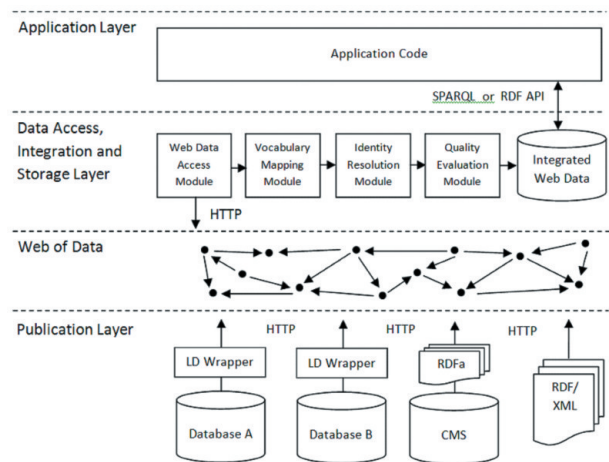


图 2 关联数据应用体系框架^[16]

以是文本/XML 等非结构化数据。

第二步:在数据访问,集成和存储层通过数据访问、术语映射等操作整合成统一的关联数据集。

第三步:应用层应用通过 SPARQL (Simple Protocol and RDF Query Language) 或者 RDF API 访问 RDF 数据。

2.2 Drupal 平台下 MOOC 关联数据的创建与发布

2.2.1 Drupal 对关联数据的支持

在 Drupal 核心的发展中 RDF 具有很长的历史。2000 年 5 月在 Drupal 诞生之初,在网站的根目录便引入了一个 RDF.PHP 的文件^[17]。该文件提供了一个基于 RDF 的面向对象 XML 解析器和一个 RDF 标题生成器。2008 年在波士顿 DrupalCon, Drupal 项目的创始人 Dries Buytaert 将语义网纳入到 Drupal 社区规划,并将其作为 Drupal 核心的一部分。2009 年的国际语义网大会上获得最佳语义网应用实践论文奖^[18]。同年,Drupal 6.X 版本上支持关联数据创建、发布的模块陆续发布,如 Resource Description Framework (RDF) 模块提供了全面的 RDF 功能和互操作性。到了 2011 年,在 Drupal 7 中 RDF 相关模块已经作为 Drupal 的核心模块发布。

2.2.2 Drupal 对关联数据互操作的支持

构建关联数据应用目的:一是遵循关联数据标准化规则生成标准 RDF 数据共享到数据云图中,二是从数据云图中消费数据,然后再将数据关联、整合到数据云图中。从关联数据提出的目的来看仅仅利于 Drupal 生成本地站点的 RDF 数据来构建 Web 应用是不够的,其最终目的是实现数据的互操作。目前,大部分 Web 上的数据无法实现数据的互操作,如果将一个网站上的数据与另一个网站上的数据合并组合,需要编写爬虫程序来抓取自己需要

的数据或者获得访问网站结构化数据的权限,然后编写相关 API 来获取相关数据。Drupal 在诞生之初便已支持对 RDF 数据的发布,上文中已有所描述。在 Web 上最常用的开放利用 RDF 数据的方式是 Web 上公开的 SPARQL 查询终端^[19]。对于 RDF 数据的开放利用,Drupal7 中也有相关模块支持。如: SPARQL 模块。

2.2.3 利用 Drupal 实现关联数据的创建与发布

在语义网络中发布 Linked Data 的准则是使用 HTTP URI 作为任何事物的标识名称^[20],而不是 ID 或者其他标识符实现唯一性识别的。

1. 使用 RDF 公开数据

RDF 是一个“实体 - 属性 - 值”(Entity-attribute-value)三元组。实体使用 URI 描述;属性使用具有通用性理解的属性名进行描述,也可以使用 URI 描述;值可以是 URI 也可以是字符串资源。如表 2 Drupal 站点中的 RDF 三元组描述示例:

表 2 Drupal 中 RDF 三元组示例

Entity	Attribute	Value
http://www.ahgf.com.cn/node/25276	Content type	25276

2. 在 Drupal 中创建自定义 RDF 映射

Drupal 中每一个内容类型都包括标题(title)、内容(Body)、创建者(Creator)、创建时间(Create Date)等字段。基于某个内容类型发布的内容,称为节点(Node)。Drupal 内容类型模型与 RDF 三元组映射关系如图 3 所示。

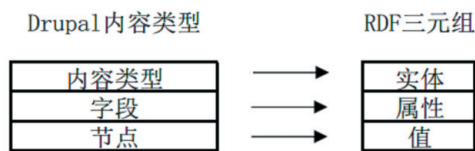


图 3 Drupal 内容类型模型与 RDF 三元组映射关系图

在 Drupal7 中使用 RDF Mapping API 可以将内容类型的任何格式的字段映射到一个 RDF 属性,任何内容类型映射到一个 RDF 类型。如,安徽国防科技职业学院网站有一个学院新闻的内容类型,则可以将其映射到类型 mo:colNews, 字段可以映射到 mo:colTitle、mo:colContent、mo:colCreateDate、mo:colDep 等。下面以创建安徽国防科技职业学院网页设计与制作 MOOC 为例说明如何在 Drupal 中创建自定义 RDF 映射。

(1) 创建模块

Drupal 中创建模块的第一步是创建.info 文件。该文件向 Drupal 系统提供模块信息及其他依赖模块信息。

```
name = webDesign RDF
description = Demonstrates an RDF mapping
using the RDF Mapping API.
```

```
core = 7.x
```

(2) 创建内容类型和字段

内容类型的创建通过实现 hook_install 钩子函数完成,该函数在模块安装时被调用。

```
// Define the node type.
```

```
$rdf_webDesgin = array(
  'type' => 'webDesgin',
  'name' => $t(' WebDesgin '),
  'base' => 'node_content',
  'description' => $t('The webDesgin node is
defined to demonstrate RDF mapping.'),
);
```

```
$content_type = node_type_set_defaults($rdf_
webDesgin);
```

```
node_type_save($content_type);
```

定义 webDesgin 字段以定义发布人字段为例

```
array(
  'web_Person' => array(
    'field_name' => 'web_Person',
    'type' => 'text',
  ),
  .....
);
```

(3) 将内容类型映射到 RDF

在.module 文件中定义内容类型、字段到 RDF 的映射。使用 hook_rdf_mapping 钩子函数完成 RDF 映射。字段映射以 MOOC 内容简介和发布人为例, webDesign 的 RDF 映射定义如下:

```
'webDesign' => array(
  'type' => 'node',
  'bundle' => ' web_Design ',
  'mapping' => array(
    'rdftype' => array('v: webDesign '),
    'title' => array(
      'predicates' => array('dc:title', 'v:name'),
    ),
    'web_Content' => array(
      'predicates' => array('v:webContent'),
    ),
    'web_Person' => array(
      'predicates' => array('v:web_Person'),
      'type' => 'rel',
```

),
.....
)
)

以有效提高教育资源的利用率。Drupal作为CMS内容管理平台,对关联数据的支持,使得预算较小的MOOC平台网站也能够将资源进行共享发布,同时,获得Web网络中的其他共享资源为己所用,为进一步解决教育资源分布不均,提高教育资源利用率作出了应有贡献。

3 结语

关联数据在MOOC教育资源共享中的应用可

注释及参考文献:

- [1] 陈大庆,丁培,叶兰. MOOC教育资源语义化关联研究[J]. 图书情报工作,2014(14):121-126.
- [2] Coursera[EB/OL].[2015-09-21]. <https://www.coursera.org/>.
- [3] edX[EB/OL].[2015-09-21]. <https://www.edx.org/course>.
- [4] 凤凰网教育[EB/OL].[2014-05-15].http://edu.ifeng.com/gundong/detail_2014_05/15/36329165_0.shtml.
- [5] 维基百科:Linked Data 词条[EB/OL].[2015-09-22]. https://en.wikipedia.org/wiki/Linked_data.
- [6] 潘有能,张悦. 关联数据研究与应用进展[J]. 情报科学,2011, (29):124-130.
- [7] The Linking Open Data cloud diagram[EB/OL].[2015-09-22]. <http://lod-cloud.net/>
- [8] State of the LOD Cloud[EB/OL].[2015-09-22]. <http://lod-cloud.net/state>.
- [9] 吴鹏飞,马凤娟. 国外关联设计教育应用项目研究与启示[J]. 电化教育研究,2013(4):114-120.
- [10] SemTech|ECS|Univerity of Sounthampton[EB/OL].[2015-09-24].<http://www.semtech.ecs.soton.ac.uk>.
- [11]MEducator Multi-type Conten Repurposing and Sharing in Medical Education[EB/OL].[2015-09-24].
<http://www.meducator.net>.
- [12]The LUCERO Project[EB/OL].[2015-09-24].<http://lucero-project.info/lb>.
- [13]黄永文,岳笑,刘建华. 关联数据应用的体系框架及构建关联数据应用的建议[J]. 现代图书情报技术.2011(9):7-13.
- [14]Talis Aspire[EB/OL].[2015-09-24].<http://www.w3.org/2001/sw/sweo/public/UseCases/Talis>.
- [15]上海市中心图书馆名录[EB/OL].[2015-09-24]. <http://data.libnet.sh.cn:8080/>.
- [16] Tom Heath,Christian Bizer. Linked Data: Evolving the Web into a Global Data Space[M/OL].[2015-09-28].
<http://linkeddatabook.com/editions/1.0/>.
- [17] Benjamin Melan?on. RDF's long lineage in Drupal[OL].[2015-09-29]. <http://data.agaric.com/rdfs-long-lineage-drupal>.
- [18] About the RDF in Drupal 7 initiative[EB/OL].[2015-09-29]. <http://semantic-drupal.com/about>.
- [19] 夏翠娟,刘炜,赵亮,等. 关联数据发布技术及其实现——以Drupal为例[J]. 中国图书馆学报.2012(1):49-57.
- [20]Berners-Lee,T.Linked Data[EB/OL].[2009-06-18].<http://www.w3.org/DesignIssues/LinkedData.html>.

Research on the Linked Data in the Sharing of MOOC Educational Resources Based on Drupal

XU Yan-dan, ZHANG Qian-jin

(Anhui Vocational College of Defense Technology, Liu'an, Anhui 237011)

Abstract: In this paper, the application of MOOC educational resources and related data in the application of educational resource sharing are studied. The solution to fragmentation of MOOC curriculum resources is proposed by using the linked data. And the realization process and the framework of the application of the linked data is introduced. Drupal—the open source CMS platform to support the linked data to make the budget of the MOOC platform for the sharing of educational resources becomes possible. Finally, the paper introduces the implementation of the Drupal to create and publish linked data.

Key words: linked data; CMS; Drupal ;MOOC

DOI:10.16104/j.cnki.xccxb.2015.04.014