

泛在学习环境下学习资源的聚合研究*

郑庆思¹ 杨现民²【通讯作者】 余胜泉³

(1. 江苏师范大学 现代教育技术中心, 江苏徐州 221004; 2. 江苏师范大学 教育研究院, 江苏徐州 221116;
3. 北京师范大学 现代教育技术研究所, 北京 100875)

【摘要】该文以泛在学习为背景,从学习资源的角度阐述了泛在学习环境下学习资源建设中存在的资源间离散程度高,资源呈现分散等问题,提出学习资源的聚合可以有效解决这些问题。该文借鉴自然语言处理及数据挖掘等相关领域的理论与方法,利用聚类分析对泛在学习环境下的学习资源进行相应的处理与计算,基于“学习元”平台,对学习资源的聚合功能进行了设计与开发,实现了基于聚类分析的学习资源聚合。

【关键词】泛在学习;学习资源;资源聚合

【中图分类号】G40-057 **【文献标识码】**A **【论文编号】**1009—8097(2013)12—0079—06 **【DOI】**10.3969/j.issn.1009-8097.2013.12.015

引言

信息通信技术的快速发展,给教育、学习领域带来了巨大的变革。泛在学习是技术变革与教育理念所催生的一种新学习方式。泛在学习以其即时性、适应性等特征为以终身学习为理念基础的学习型社会的构建提供了有力支持。^[1]泛在学习是一种随时随地都可能发生的学习形式,面对的学习者群体也是极为庞大、复杂的,这要求必须有海量的,符合学习者个性的学习资源来满足学习者的学习需求。要满足泛在学习对学习资源的需求,需要采取“群建共享”的开放、协同的资源建设模式,实现泛在学习资源的持续性生成和可进化发展,以此来满足学习者学习的各种个性化需求^[2]。

但学习资源“群建共享”的建设模式也会产生这样一个问题:这些由学习者生产的资源的离散程度高,即学习资源之间是没有关系的或可用关系很少,缺少传统资源建设模式下的严格的学习资源学科分类体系。杨丽娜等^[3]从泛在学习虚拟社区的资源建设的角度分析了目前资源建设存在的资源组织的非结构化、资源“生长”的无序性以及资源关系的孤立化等问题。通过对学习资源的聚合,可以有效解决上述问题。同时,学习资源的聚合可以为学习者在发现资源使用资源的过程中提供一定的帮助。本研究基于学习元平台^[4]提出一种基于聚类分析的学习资源聚合方法,在该平台上检验该方法的聚合效果并借助第三方插件实现资源聚合的可视化展现。

一 学习资源聚合研究发展现状

目前,学习资源聚合的技术手段主要有 RSS (Really Simple Syndication, RSS)、Mashup 技术,数据挖掘技术以及新兴的语义 Web (Semantic Web) 技术。RSS 技术最早应用于教育领域。国内关于 RSS 技术应用于教育领域的探索早在

2004 年便已经开始,胡三华,汪晓东^[5]在研究博客在教育领域的价值时,提到了 RSS 技术,并尝试将其应用于教育领域。陆峰^[6]在研究教育资源库的建设时提出利用 RSS 技术向用户进行资源的推荐。张力^[7]对 RSS 技术在聚合个性化的网络教育资源的可能性进行了探讨。与此同时,也有研究者将 RSS 技术的应用视角放在远程教育之中。^[8]国外 RSS 技术在教育中应用与研究的发展脉络与国内相似,对该技术的研究始于学术播客,RSS 技术的作用是集成会议视频的链接与相关的文件^[9]。有研究者开始研究 RSS 技术在分布式教育资源聚合与个性化方面发挥的作用进行了研究^[10]。但是,RSS 只是简单地将信息、资源进行了集成。Mashup 通过对不同的数据源之间进行复合,则可以产生出很多附加价值,且 Mashup 技术更符合 Web2.0 的特点,因而 RSS 技术在很多方面逐渐被 Mashup 技术取代。

Mashup 技术出现于 2005 年,在对其从教育视角进行研究之初,很多研究者便将重点放在了教育资源的聚合上。有研究者认为,Mashup 技术可以“为资源的深度整合提供途径,使既有资源体现出新的价值”^[11]。在教育领域的研究者熟悉开发流程之后,开始有研究者利用 Mashup 技术解决非正式学习的资源整合问题^[12]。Mashup 需要多个数据源以完成对信息、资源的整合,这就要求数据源在格式上的统一,且能够相互识别;多个数据源也可能会带来一些数据安全及版权上的问题。

已有研究者利用聚类分析手段对教育教学中的资源进行处理。高少琛^[13]对资源检索结果进行聚类分析处理的优化。资源检索是将符合用户输入信息的所有资源集中展现给用户,因此我们认为通过搜索引擎对资源进行检索本身就是一种基于关键字的隐性的资源聚合行为。也有研究者利用聚类分析技术对个性化推荐及信息服务技术进行优化与改进。杨

涛^[14]利用聚类分析技术在基于用户兴趣模型的个性化学习环境中更好地为用户推荐学习内容；孙守义^[15]利用聚类对传统个性化信息服务技术进行改进。

语义 Web 是在其基础之上将信息设计成机器也可读，是对传统互联网的扩展，互联网中的信息被额外赋予了明确的、机器可读的含义，使人与机器之间可以更好地交互与协作。语义 Web 技术在 e-Learning 领域资源的组织、重用和共享等方面对学习资源聚合的实现思路提供很好的指导。

以上几种资源聚合的技术手段各有特点，RSS 技术目前已使用较少，多用于简单、线性的信息推荐；Mashup 技术方兴未艾，这种技术在糅合多个网站的信息然后推送给学习者上具有较大的优势；聚类分析则较常用于资源检索结果的处理以及个性化资源推荐上；语义 Web 技术则具有在技术上前瞻性的优势。目前，学习元平台内已实现的基于语义的学习资源聚合的方法对学习资源之间的关联即学习资源间的语义关系依赖程度较高^[16]，生成效率较低，因此本研究将采用聚类分析的手段来快速高效的实现学习资源的聚合。

二 资源聚合的设计及实现

1 总体框架设计

在本研究的整个研究过程中，首先需要对学习资源进行初级聚合。学习资源的初级聚合是指，对最基本的资源实体进行聚合，是基于数据挖掘领域中聚类分析的相关技术与手段进行。聚类分析的结果是将数据某种属性或其本身按照其相似性聚集在一起。利用聚类分析这种手段进行学习资源的聚合可以有效地将相似及相关的学习资源聚合在一起。如图 1 所示为基于聚类分析的学习资源聚合整体框架图。在整体的设计中既包括了学习资源聚合的设计，也包括了对聚合效果的检验及对资源聚合效果的可视化实现。

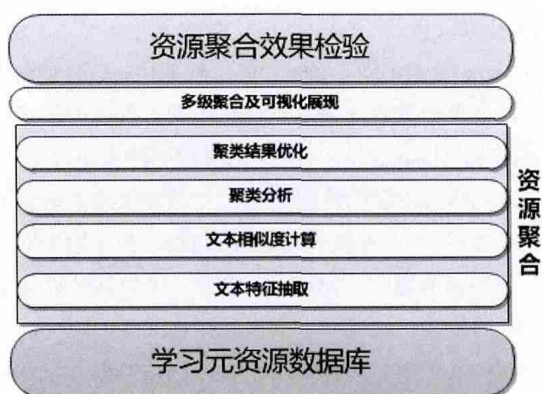


图 1 资源聚合整体框架图

2 基于聚类分析的学习资源聚合

聚类分析是指按照数据某种属性的相似程度把未标明类别的数据集合划分成自然类簇的数据分析过程。其基本要求是：每个类簇中的数据之间的相似程度尽可能的高，不同类簇之间数据的相似程度尽可能小。其基本步骤如下：

(1) 对数据集进行表示和预处理，包括数据清洗、特征选择或特征抽取；(2) 给定数据之间的相似度或相异度机器定义方法；(3) 根据相似度，对数据进行划分，及聚类；(4) 对聚类结果进行评估。

聚类分析是数据挖掘中重要的技术手段。在本研究所依托的平台中，大多数学习资源是以文本类的资源，本研究即是利用聚类分析可以发现具有相似主题的文本的特性，分析学习资源，以发现相似的学习资源，并将其聚集成类。

(1) 数据准备与处理

在利用聚类分析对学习资源进行聚合之前，需要对学习资源中的文本信息进行一系列的处理，以便聚类分析的进行。这些处理包括对学习资源的结构的分析，文本内容的数据清洗以及分词等等。

学习资源的结构分析：在按照学习资源结构化程度可以分成结构化资源，半结构化资源，非结构化资源。结构化资源均有确定的逻辑结构，在结构中每个部分均有确定的含义与清晰的语义信息或逻辑信息描述，仅供开发人员使用。半结构化资源具有一定结构，在存储时有明确的标题、标签、简介等元数据信息，但语义信息及逻辑结构信息不够确定，如 HTML 网页中，有些字段的含义是确定的（如 title 中的内容），而有些则是不确定的（如 table 中的内容）。非结构化资源的结构较为杂乱，没有明确的信息，很难按照某个概念进行抽取。结构化、半结构化的资源经过一定处理后便可进入下一阶段的分析，而非结构化的资源处理起来相对复杂与麻烦。

本研究依托的平台中，绝大多数资源均为半结构化资源并含有少量语义信息，因此可以使用这些信息进行有效、快速地进行聚合。同时，鉴于平台中的用户多为教师以及有较高教育背景的人群且平台具有的智能审核功能。因此，我们认为这些用户创建资源时填写的标题、标签等内容与资源本身的内容是高度相符的。这是利用这些信息进行聚合的一个大的前提。

数据清洗：不同于数据库中的“脏数据”清洗，这一环节的主要任务是在聚类过程不用到或会对聚类过程产生干扰的内容清除出去。主要是将学习资源本身所带的一些特殊标记或符号如 html 格式信息等一类内容进行清洗。

文本分词：分词是自然语言处理的基础，在对学习资源进行聚类分析的过程中，经过数据清理后要进行的就是对学习资源的文本内容进行分词。对文本进行分词后，再经过一些处理即可达到机器自动识别语义的效果。我们可以很好地理解哪些字组成在一起是一个词语，但对计算机而言则则需要分词算法才能较好的分辨。

目前的分词算法按照其分词原理可分为三类：基于字符串匹配的分词方法、基于理解的分词方法、基于统计的分词方法。按照是否与词性标注过程相结合，又可以分为单纯分词方法和分词与标注相结合的一体化方法。

目前平台中用到的分词器有 IKAnalyzer 分词组件及 ICTCLAS 1.0 分词组件。IKAnalyzer 是结合词典分词和文法分析算法，ICTCLAS 1.0 除了包含中文分词的功能外还有词性标注；命名实体识别；新词识别等几个功能。经过综合对比，最终选择 IKAnalyzer 作为本研究中的分词器。但 IKAnalyzer 自带词典所涵盖的词语有限，且没有涵盖专业名词与平台中学习资源带有的一些专有名词，因此我们建立了自定义用户词典。自定义用户词典主要是通过人工筛选及自动加载资源分类信息两种方式进行。

(2) 文本特征抽取

文本特征抽取的主要功能是在不损伤文本核心信息的情况下尽量减少要处理的单词数，以此来降低向量空间维数，从而简化计算，提高文本处理的速度和效率。文本特征选择对文本内容的过滤和分类、聚类处理、自动摘要以及用户兴趣模式发现、知识发现等有关方面的研究都有非常重要的影响。特征抽取的方法有很多，本研究中使用的是 TFIDF 函数：

$$\psi = tf_i(d) * \log\left(\frac{N}{n_i}\right)$$

其中 N 为所有文本集中文本的数量， n_i 为含有词条 t_i 的文档数目。

TFIDF 函数在这一类应用中是使用较多的一种方法，其基本原理是：如果某个词条在某一文本中出现的频率 TF (Term Frequency) 高，并很少出现在其他文本中，则认为此词条具有较好的区分能力；IDF (Inverse Document Frequency) 是反文档频率，其基本原理是：如果包含词条 t 的文本数量越少，则 IDF 越大，同样说明此词条 t 具有较好的区分能力。如果某一类文本集 C 中包含词条 t 的文档数为 m ，而其他类包含 t 的文本总数为 k ，则所有包含 t 的文档数 $n=m+k$ ，当 m 大的时候， n 也大，根据 IDF 公式：

$$idf = \log\left(\frac{N}{n_i}\right)$$

得到的 IDF 的值会小，说明该词条 t 类别区分能力不强。但存在这样一个问题，如果一个词条在一类文档中频繁出现，则说明该词条能够很好代表这个类的文本的特征，这样的词条应该给它们赋予较高的权重，并作为该类文本的特征词以区别于其它类文档。这便是其不足之处了。事实上，IDF 是对 TF 进行噪音抑制的加权，并且认为文本频数小的词条区别度越高，文本频数大的词条区别度越小，这并不完全正确。但 TFIDF 算法相对简单、原理清晰，且拥有较高的准确率与召回率，因此在本研究中我们选择它对分词后的文本进行加权以构建向量空间。同时，我们在 TFIDF 算法的基础上进行了改进：对一些特殊词汇进行额外加权以获得更好的聚合结果。

依托学习元平台我们实现了 TFIDF 对文本的加权，算法流程如下：

名称: TFIDF 加权算法
输入: 分词之后的文本
输出: 带有权重词条列表
关键步骤:
Step1 输入分词之后的文本
Step2 统计词频并过滤相同词条
Step3 统计词条在该文本集中出现的频率
Step3 计算词条逆文本频度
Step5 计算词条的权重
Step6 匹配关键词汇进行额外加权
Step7 重复 Step2 直至结束

(3) 文本相似度计算

数据间的相似性是聚类的核心。在数据挖掘领域中，一般有两种方式表示数据间的相似程度。一种是把数据看成 V 维空间(v 为数据集中数据的个数，下同)中的点，用点之间的距离表示样本之间的相似性程度，称为距离系数，以 d 表示； d 越小则表示相似程度越高， d 越大则表示相似程度越低。另一种是用相似系数来度量数据之间的相似程度，相似系数越大则相似程度越高，相似系数越小则表示相似程度越低。相似系数体现数据间的相似程度，反映数据间相对于其某些属性的相似程度。不同类型的数据需要用不同相似性度量方式才能获取较好的聚类效果。对于数值型的数据，两个数据的相似程度是指它们在欧氏空间中的互相邻近的程度；而对分类型的数据而言，两个对象的相似程度是与它们取值相同的某种属性的个数有关。本研究通过试验最终选择余弦相似度作为相似系数作为不同学习资源间的相似性度量。

文本在进行计算之前，需要用数学方法对其表示。文本表示模型一般有布尔模型、向量空间模型及概率模型等。文本表示模型的选择是根据实际需要来选择的。在本研究中，考虑到使用余弦相似度作为相似性度量，因而使用向量空间模型来表示文本。向量空间模型 (VSM: Vector Space Model) 是这样一种方法，此模型假设词与词之间不相关，用向量表征文本，简化文本中的词之间的关系，使文本可以用十分简单的向量表示。向量空间模型将文本集看作是由文本均被表示为其中的一个泛化特征向量 $V(d) = (t_1, \omega_1(d)); \dots; (t_i, \omega_i(d)); \dots; (t_n, \omega_n(d))$ 组成的向量空间 V ，其中 t_i 为词条项， $\omega_i(d)$ 为 t_i 在文本集 d 中的权值， n 为文本集中文本的数量。^[17]将 d 中出现的所有单词作为 t_i ，与要求 t_i 是 d 中出现的所有短语，可以文本表示提高准确性。 $\omega_i(d)$ 一般被定义为 t_i 在 d 中出现的频率 $tf_i(d)$ 的函数，即 $\omega_i(d) = \psi(tf_i(d))$ 。常用的频率函数 Ψ 有布尔函数：

$$\psi = \begin{cases} 1, & tf_i(d) > 1 \\ 0, & tf_i(d) = 0 \end{cases}$$

文中提到的 TFIDF 函数。

如此，两个文本的相似度可以在其在向量空间中的特征向量的夹角的余弦值来度量。其公式如下：

$$Sim(d_i, d_j) = \cos \theta = \frac{|\sum_{k=1}^n W_{ik} * W_{jk}|}{\sqrt{(\sum_{k=1}^n W_{ik}^2) (\sum_{k=1}^n W_{jk}^2)}}$$

其中， W_{ik} 、 W_{jk} 分别表示文本 d_i 和 d_j 第 k 个特征项的权值， $1 \leq k \leq N$ 。

(4) 聚类分析算法

到目前为止，人们提出了许多聚类算法，包括：划分方法、层次方法、基于密度的方法、基于网格方法、基于模型方法、模糊聚类、一群聚类方法、高维数据聚类方法等等。每种聚类方法都有各自的特点，在这里我们选择一种较为常用的聚类算法来对学习资源进行聚类

k-means 算法^{[19][20]}是一种典型的基于划分的方法，它在 1965 年即被人提出，尽管年代久远，但仍然是十分受欢迎算法，被认为是迄今为止在科学和工业领域中应用最为广泛的聚类算法^[21]。这是一种迭代的聚类算法，在迭代的过程中，不断的移动类簇的对象直到类簇的平均值达到回归要求，每个类簇亦使用其对象的平均值来表示。**k-means** 的算法复杂度上界为 $O(k*t*n)$ 其中 t 为迭代次数， n 为文档个数， k 为类别个数。依据上述表述，**k-means** 算法描述如下：

名称: k-means 算法

输入: 簇的数目 k ，包含 n 个文本的特征向量

输出: k 个簇，使平方误差准则最小

关键步骤:

Step1 输入文本特征权值向量

Step2 任意选择 k 个向量作为初始的簇中心

Step3 对每个点确定其聚类中心点

Step4 根据簇中对象的平均值，将每个对象重新分配

Step5 更新簇的平均值

Step6 转到 **Step4**，重新计算每个簇的平均值。这个过程不断重复直到某个准则函数不再明显变化或聚类的对象不再明显变化为止。

一般而言，**k-means** 算法的准则函数采用平方误差准则，

定义为
$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2$$

其中， E 是数据集中所有对象与相应聚类中心的均方差之和， p 为给定的数据对象， m_i 为聚类 C_i 的均值。

k-means 具有高效率，并有效处理大文本集的优点。但该方法需要预先指定 k 值和初始划分，从而容易使聚类结果受到影响，这是它的最大的缺点。

(5) 聚类结果优化

由于聚类分析算法及实际数据本身的一些不能避免的缺陷，聚类分析的结果一般与样本数据有很大的差异。比如，在提供 1000 个结构完整，信息完备的样本数据，并确定为 33 个类别，聚类的准确率可能达到 90% 以上。但实际中的数据可能会有各种缺陷，而且类别的个数是不确定的，聚类分析

的结果可能会比较差，这样的结果是没有办法用于资源聚合的。因此，需要到聚类分析的结果进行程序上的优化，以保证资源聚合可以有效地完成。

本研究通过对聚类分析结果中较大的类簇（类簇中资源数量大于 100）及较小的类簇进行分别优化。聚类分析后，类簇中资源数量大于 100 的 8 个类。**ClusterInfo** 类的功能设计为存储类簇基本信息，其中 **clusterId** 为类簇的标识，**lcnnumber** 字段为该类簇中学习资源的数量，**keywords** 为该类簇中出现大于 1 次的关键词。

此外，类簇中学习资源数量较少的类簇也有很多，由此可知，对聚类分析的结果进行再优化是十分必要的。对资源数量大于 100 的类簇采用再次聚类的方式进行优化，直接调用原有的聚类分析方法即可；而资源数量较小的类簇的优化方式是将其打散，与其他资源数量适中的类簇进行形似度匹配，并将资源归类到与其相似度较高的类簇中。同时，在资源集合中经常会有噪声数据，没有办法归到任何一类当中，由于这种资源数量不是很多，所以其处理方式是集中放在一个类簇中，后期再具体分析并归类，或由人工判断具体放在哪一类中比较合适。

经过优化，类簇中的资源数量明显变得更加均衡，首先消除了那些只有 1~2 个资源的类簇，其次将噪声数据都集中到了一个类簇中，便于后续优化处理的进行。最终得到相对合理有效的资源聚合的结果。

(6) 聚合结果可视化实现

数据可视化是关于数据之视觉表现形式的研究。聚合结果的可视化导航，采用 **datav.js** 可视化组件 (<http://datavlab.org/>)。经过比较，本研究采用 **treemap** 视图对资源进行可视化。**treemap** 数据输入的格式可以是二维数组，其下再用二维数组嵌套，最终实现层级的展示。为此我们根据第一次资源聚合的结果用同样的方法进行了二、三级的聚合。数据的基本结构如下：

```
{ "name": "根目录", "children": [ { "name": "三级聚合", "children": [ { "name": "二级聚合", "children": [ { "name": "初级聚合", "children": [ { "name": "原始资源", "size": 1 } ] } ] } ] } ] }
```

通过引入 **datav.js** 文件，创建接入代码接收数据即可完成资源的可视化。将资源聚合的结果用于可视化导航，效果如图 2 所示。最顶层显示的是三级聚合的结果，共 22 个类簇，各区块的大小由该类簇中所包含的学习资源总数决定。

以图 2 中最大的区块为例，点击该区域后进入下一级，在第二级中，在鼠标悬浮区域将显示该类簇的关键词信息以及该类簇中所包含的资源数量，同时该区域高亮显示。鼠标右键单击时则返回上一级。这种可视化的方法可以用户快速了解系统内有哪些资源，通过对关键词的追踪与点击相应区域，可以获知该关键词相关的资源有哪些。在后续的工作中，可以设计与实现相应的功能，使学习者可以直接在此处获取资源的整体内容。



图2 可视化效果截图

三 聚合效果检验

资源聚合的效果检验主要从聚合所生成的类簇的质量与数量两个方面来检验。经过运行资源聚合方法，共生成类簇444个，而原系统内语义聚合的方式共生成知识群的数量仅为68个，新的聚合方式生成的类簇的数量是现有自动聚合方式生成知识群数量的6.53倍。即便排除聚合质量不佳的一些类簇，其聚合的效率还是较原有方式高很多。图3为聚合结果页面截图。

截止到2012年11月1日，平台内通过审核且开放浏览的知识群共776个，其中由用户创建的知识群708个，自动聚合生成68个。通过动态语义聚合方法生成的知识群仅占系统内知识群总数的8.76%，若将新生成的类簇全部转换成知识群，则自动生成的知识群数量将占系统内知识群总数的36.4%。这是一个很大的提高，可以使用户更方便的获取到资源。

我们对资源聚合的类簇取其总数的10%用人工的方式进行了抽样检查，计44个类簇，发现在44个类簇中，有31个类簇的聚合质量较高，聚合准确率达到70.45%。虽然较语义聚合方式的准确率一些，但考虑到类簇总量大，聚合效率

高，总体而言聚合的效果是可以接受的。

通过调整资源聚合的相关阈值，将产生不同的聚合结果。如相似度阈值越高，单个类簇的聚合质量便会越好等等，在后期的工作中可以调整阈值来满足系统的实际需求。经过聚合，可以有效的将相似的资源聚合在一起并生成类簇。但一些类簇所含资源数量较多，类簇内相似度相对较低。

本研究的主要内容是学习资源的聚合。资源聚合本身的目的之一便是方便学习者获取资源，因此在资源聚合完成后需要有合理的呈现方式。

四 总结与展望

利用数据挖掘领域中的聚类分析技术，完成了面向资源的聚合，简化学习者获取资源的过程，减少学习者获取资源的时间，提高学习者的效率。本研究在资源聚合的效果上还算理想，但还有许多地方有待改进。可视化导航的应用目前主要在数据分析领域，主要用于数值型数据的可视化展示，对于学习资源的可视化导航等功能研究不多。在本研究中，我们主要利用层级目录式的可视化完成对资源的可视化导航，后续工作可以在已有的可视化基础上做更加深入的研究。

学习元聚合

学习元名称	创建者	类别	主题词
优秀信息技术教师特质分析与教学策略	林禹江	15	优秀 信息技术教师 特质 分析 教学策略
优秀信息技术教师的特质分析 by 邵辉	邵辉	15	优秀 信息技术教师 特质 分析 by
信息技术教师所需要的素质	姚英	15	信息技术教师 需要 素质
优秀信息技术教师之我见	瞿超	15	优秀 信息技术教师 之我见
优秀信息技术教师特质之我见	赵智超	15	优秀 信息技术教师 特质 之我见
信息技术教师优秀特质之我见	韦双舒	15	信息技术教师 优秀 特质 之我见
优秀信息技术老师之特征分析	覃信服	15	优秀 信息技术 老师 特征分析
优秀信息技术教师特质之我见-石胜兵	石胜兵	15	优秀 信息技术教师 特质 之我见
优秀信息技术教师特征分析	周思博_josie	15	优秀 信息技术教师 特征分析
例说信息技术课堂中常用策略	瞿超	19	信息技术课堂 常用 策略

总共 5228, 当前 151 - 160 首页 | 上一页 | 下一页 | 末页 16/523 Page 16 GO

图3 聚合结果页面截图

参考文献

- [1]杨现民,余胜泉.生态学视角下的泛在学习环境设计[J].教育研究, 2013, (3): 103-110.
- [2]余胜泉,杨现民,程罡.泛在学习环境中的学习资源设计与共享——“学习元”的理念与结构[J].开放教育研究, 2009, (1):47-53.
- [3]杨丽娜,肖克曦,刘淑霞.面向泛在学习环境的个性化资源服务框架[J].中国电化教育,2012,(7):84-88.
- [4]杨现民,余胜泉.学习元平台的设计开发及其应用场景分析.电化教育研究, 2013, (3):55-61.
- [5]胡三华,汪晓东.博客在教育教学中的应用初探[J].远程教育杂志, 2004, (1): 10-12.
- [6]陆峰.RSS 技术在教育资源库建设中的应用[J]. 教育信息化,2006,(16):50-53.
- [7]张力.对 RSS 聚合个性化网络教育资源的探讨[J].中国医学教育技术,2006,(6):512-515.
- [8]张伟,高宏卿. RSS 技术在网络远程教育中的应用[J]. 教育信息化,2006,(17):76-77.
- [9]Tripp, Jacob S; Duvall, Scott L; Cowan, Derek L; Kamaau, Aaron W C, etc. Academic podcasting: quality media delivery[J]. AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium, 2006: 1125
- [10]范剑文.分布式教育资源聚合与个性化应用研究[J].电化教育研究,2009,(9):80-83.
- [11]马小强.Mashup 的教育价值及启示[J].电化教育研究, 2008, (9):56-60.
- [12]王秀芝,邹霞.利用搜索 Mashup 技术解决非正式学习资源整合问题[J].现代教育技术,2010,(2):116-119.
- [13]高少琛.基于查询拓展和聚类技术的资源检索系统的研究与应用[D].上海:上海交通大学,2007.
- [14]杨涛.基于 Web 的个性化学习系统的研究[D]北京:中国石油大学,2007.
- [15]孙守义.基于聚类的教育资源个性化信息服务[D]南京:南京师范大学,2008.
- [16]杨现民,余胜泉,张芳.学习资源动态语义关联的设计与实现[J].中国电化教育, 2013, (1): 70-75.
- [17]G. Salton , A. Wong , C. S. Yang, A vector space model for automatic indexing[J]. Communications of the ACM, 1975,(17):613-620.
- [18]Forgy, E.W. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. In: Biometric Society Meeting, Riverside, California, Abstract in Biometrics, 1965:768.

Research on Learning Resource Aggregation in Ubiquitous Learning Environment

ZHENG Qing-si¹ YANG Xian-min² YU Sheng-quan³

(1. Modern Educational Technology Institute, Xuzhou Medical College, Xuzhou 221004, China;

2. Department of Educational Science, Jiangsu Normal University, Xuzhou 221116, China;

3. Institute of Modern Educational Technology, Beijing Normal University, Beijing 100875, China)

Abstract: This paper based on the environment of ubiquitous learning, we describe the problem of the learning resource which was loosely and the structured organization. We thought the aggregation of the learning resource could resolve these problems. This paper referenced the theory and method of the nature language process and data mining, used the cluster analyzer method to process and calculates the learning resource in the ubiquitous learning environment. Based on the “learning cell” platform, design and implement the function of learning resource aggregation.

Keywords: ubiquitous learning; learning resource; resource aggregation

*基金项目：本文系国家自然科学基金项目“泛在学习的资源组织模型及其关键技术研究”(项目批准号：61073100)和江苏师范大学优秀博士人才引进科研支持项目“网络课程资源的动态生成模式研究”(项目批准号：12XWR015)研究成果之一。

作者简介：杨现民，江苏师范大学教育研究院，研究方向为移动与泛在学习、数字化学习资源设计。

收稿时间：2013年4月13日

编辑：李婷