一种基于组合策略的聚类描述方法及其应用

章成志[1,2]

(1. 中国科学技术信息研究所，北京 100038；2. 南京理工大学 信息管理系，江苏 南京 210094)

摘 要：针对 DCF 聚类描述法存在的问题，提出一种基于组合策略的聚类描述方法，即综合利用“先描述、后聚类”和“先聚类、后描述”的优点，解决聚类描述的可理解性问题。实验结果表明该方法的有效性，将该方法用于搜索结果聚类这一应用中。

关键词：聚类描述；DCF；文本聚类；搜索结果聚类

中图分类号：G354；TP391　　**文献标识码**：A　　**文章编号**：1007-7634(2009)07-1079-06

Method and its Application of Document Clustering Description Based on Combination Strategy

ZHANG Cheng-zhi[1,2]

(1. *Institute of Scientific & Technical Information of China*, *Beijing* 100038, *China*; 2. *Department of Information Management*, *Nanjing University of Science & Technology*, *Nanjing* 210094, *China*)

Abstract: The DCF (Description Comes First) method can generate document clustering description. A method based on combination strategy, i.e. combination of the DCF and DCL (Description Comes Last) is proposed to solve the problem of the weak readability of clustering description in this paper. Experimental results show that the method is effective, and the method is used to describe the search result clustering.

Key words: clustering description; DCF; document clustering; search result clustering

1 引 言

标注文档集合聚类后生成的类簇，是文本聚类应用中不可或缺的一项任务[1]。标注文档集合聚类后生成的类簇，可以让用户更容易通过类簇的标签来了解各个类簇的主题，节省信息浏览时间。通常，该任务被称为文本聚类结果的类簇标注(Cluster Labeling)[2-5]、类别标注[5]、类簇命名(Cluster Naming)[6-7]、标签识别(Label Identification)[5,8]、主题发现或识别[5]、文本聚类结果描述（简称聚类描述，Cluster Description，如无特殊说明，本文后面提到的聚类描述就是特指文本结果聚类描述）[9]、描述聚类(Descriptive Clustering)[9]或文本聚类结果类别标题的自动生成(Title Generation for Clustered Documents)[1]。在机器学习和数据挖掘领域，聚类描述是概念聚类的后处理部分[10]。

现有的文本聚类方法中共存的问题，也是查询结果聚类(Search Results Clustering)的难点问题之一，就是聚类结果的有效描述问题。传统的聚类算法直接用于文本聚类上，存在的突出问题就是算法的有效性问题，因为传统的聚类算法只对对象进行聚

收稿日期：2009-06-01

基金项目：“十一五”国家科技支撑计划重点项目(2006BAH03B02)；中国博士后科学基金资助项目(20080430463)；南京理工大学科研启动基金项目(AB41123)

作者简介：章成志(1977-)，男，安徽人，讲师，博士后，从事信息组织、信息检索、数据挖掘及自然语言处理研究.

类，不负责对聚类后生成的类簇进行概念描述和解释。因此，必须针对文本聚类的特别要求，探寻专门解决文本聚类描述这一问题的方法。

聚类描述是帮助用户迅速确认生成的文档类相关与否的重要信息。聚类描述是一项很具有挑战性的工作[8]，具有重要的研究意义和应用价值。按照聚类描述生成的自动化程度，可以将其分为人工描述方法和自动描述方法。

Patrick、Wolfgang、Lai、Wu 等人通过人工描述方法完成聚类描述工作[11-12]。

自动化的聚类描述，主要从聚类生成的类簇中提取重要的词语，根据聚类算法的不同，相应的词语重要性计算方法也有所不同[1]。1992 年，Cutting，Karger 和 Pedersen 等人在 Scatter/Gather 系统中，利用归一化的词语频率作为词语权重，选择权重较大的词语组成列表作为类别描述[13-14]。后来，Muller 和 Dorre 等人则直接将聚类类簇的前 N 个最高频次的词语作为聚类描述[15]。1996 年，Anton 和 Croft 等人则用关键短语来进行类别描述[16]。1998 年，Zamir 和 Etzioni 将类簇的文档集中的出现频次(TF)高的最长短语作为聚类描述[17]。2001 年，Lawrie 和 Croft 等人从聚类生成类簇的成员中提取 TF*IDF 值大的词语作为聚类描述[6]。2002 年，Glover 和 Pennock 等人通过对文档集合中的父子、同类特征建立统计模型，通过层次关系进行推理得到聚类描述[18]。2006 年，Tseng & Lin 等人借助了 WordNet 作为外部资源提取类别词作为聚类描述符，并利用上位词搜索算法将类别描述符转换为宽泛的、可作为聚类描述的词语[1]。同年，Pucktada 和 Jamie 综合利用描述词在类簇本身、父类簇和文本集合的统计信息得到类簇标注[3]。

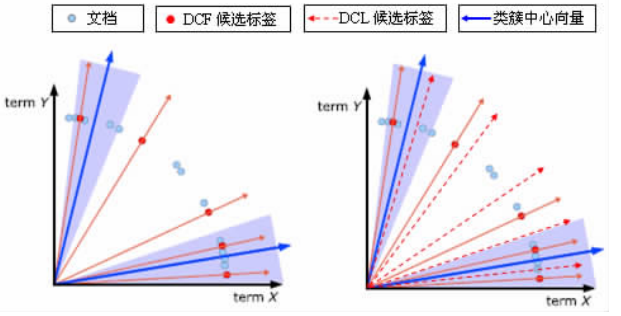
2006 年 Dawid 提出了一种全新的方法，即 DCF(Description Comes First，聚类描述先出现)算法，用它来解决传统方法中存在的聚类描述可读性不强的问题，与传统的聚类结果描述算法不同的是，DCF 算法是在文本聚类完成的同时，聚类标签也生成出来。该方法生成的聚类标签其实是类簇中心向量的替代物[9]。由于 Dawid Weiss 是采用事先生成标签来代替聚类后生成的类簇中心向量，这样一方面使得聚类描述与聚类类簇的中心向量之间存在一定的“语义间隔”，另一方面，这与人们“先聚类、后描述”的直觉相违背，减低了聚类描述的可解释性。本文拟综合利用“先描述、后聚类”和“先聚类、后描述”的优点，解决聚类描述的可理解性问题。

2 基于 DCF-DCL 组合策略的聚类描述方法

为了进行对照实验，本文首先分别给出了基于 DCF、DCL 的描述方法。其中 DCF 方法采用 Dawid 聚类描述方法[9]。

2.1 几种对照的聚类描述方法

为了对基于 DCF-DCL 组合策略的聚类描述算法(记为 DCF-DCL)进行性能分析，本文增加了两种对照的聚类描述方法作为基准模型，即 DCF 方法的聚类描述方法(记为 DCF)与 DCL 方法。其中，DCL 中根据描述词的生成方法的不同进一步分为基于中心向量的描述方法(记为 DCL1)与基于 SVM 的描述方法[19](记为 DCL2)。



(a) (b)

图 1 DCF[9](a)与 DCF-DCL 聚类描述方法(b)示意图

(1)DCF 方法。图 1(a)给出了 DCF 方法生成聚类描述的过程。2006 年 Dawid Weiss 首先提出 DCF[9]方法，他利用了类簇中心向量与事先生成的聚类标签进行了比较，增强了聚类标签的可信度。具体的算法描述参见文献【9】。

(2)DCL1 方法。这里的基准模型 DCF1 是利用聚类类簇的中心向量作为聚类描述词的来源。本文取概念向量中权重前 5 位的词语作为聚类描述。

(3)DCL2 方法。本节将基于 SVM 聚类描述模型作为第二种 DCL 基准模型。关于 SVM 聚类描述模型详细说明可参见文献【19】。

2.2 基于 DCF-DCL 组合策略聚类描述算法

如前所述，Dawid Weiss 首先提出 DCF[9]方法一方面增强了聚类标签的可信度，另一方面该方法存在语义间隔问题(即聚类描述与聚类类簇的中心向量之间存在一定的“语义间隔”)与聚类描述直观解释问题(即 DCF 方法与人们“先聚类、后描述”的直

觉相违背,减低了聚类描述的可解释性)。为了利用DCF聚类描述方法的优势,并且克服DCF中存在的语义间隔与聚类描述直观解释问题,本文提出基于DCF-DCL组合策略的聚类描述算法。

图1(b)给出了DCF-DCL方法生成聚类描述的过程。该方法综合考虑DCF与DCL方法生成的聚类描述。

算法:DCF-DCL聚类描述算法.

输入:文档集合 $D=\{\vec{d}_1,\cdots,\vec{d}_j,\cdots,\vec{d}_n\}$;

文档聚类参数:文档聚类的类簇数目K;

聚类描述参数:DCF与DCL聚类描述权重系数分别为α、β,且$\alpha+\beta=1$;

输出:K个类簇C_i及其类簇描述$DC_i, i\in[1,K]$;

步骤:

//Setp1: DCF聚类描述预处理:高频关键词提取

Step1.1: 利用CRF自动标引模型对文档集合D中的每篇文档,$\vec{d}_j$进行关键词抽取,生成,$\vec{d}_j$的关键词集合Keyword_j,得到D的关键词集合$\text{Keyword}=\{\text{Keyword}_1,\cdots,\text{Keyword}_j,\cdots,\text{Keyword}_n\}$;

Step1.2: 对关键词集合Keyword进行频次统计,得到频次最高前M个关键词组成集合$\text{Keyword-TopM}=\{\text{Keyword}^1,\cdots,\text{Keyword}^i,\cdots,\text{Keyword}^M\}$,;

//Step2:文本聚类

Step2.1: 利用样本加权聚类算法对文档集合进行聚类,生成D的一个划分为$C=\{C|\ C\subseteq D\}, D=\cup_{Ci\in C}, C_i, i\in[1,K]$,得到每个类簇的中心向量$CV_i, i\in[1,K]$;

//Step3: DCF聚类描述生成

Step3.1: 计算Keyword-TopM与每个类簇的中心向量CV_i的相似度;

Step3.2: 将与CV_i最相似的前P个关键词Keyword^j($j\in[1,M]$)作为类簇$C_i(i\in[1,K])$的聚类描述,得到类簇C_i的DCF聚类描述$\text{Keyword-DCF}=\{\text{Keyword-DCF}^1,\text{Score (Keyword-DCF}^1),\cdots,\text{Keyword-DCF}^i,\text{Score (Keyword-DCF}^i),\cdots,\text{Keyword -DCF}^P,\text{Score(Keyword-DCF}^P)\}, i\in[1,P]$,其中,$\text{Score (Keyword-DCF}^i)$为$\text{Keyword-DCF}^i$与$\text{CV}_i$的经归一化后的相似度;

//Step4: DCL聚类描述生成

Step4.1: 利用DCL1或DCL2模型生成每个类簇$C_i(i\in[1,K])$的前Q个DCL聚类描述$\text{Keyword-DCL}=\{\text{Keyword-DCL}^1,\text{Score (Keyword-DCL}^1),\cdots,\text{Keyword-DCL}^i,\text{Score(Keyword-DCL}^i),\cdots,\text{Keyword -DCL}^Q,\text{Score(Keyword-DCL}^Q)\}, i\in[1,Q]$,其中,$\text{Score(Keyword-DCL}^i)$为$\text{Keyword-DCL}^i$的权重(DCL1中)或以标注得分形式得到的分值(DCL2中)的归一化数值;

//Step5: DCF-DCL聚类描述生成

Step5.1: 对于每个类簇$\text{C}_i(i\in[1,K])$的Keyword-DCF和Keyword-DCL候选描述词集合中的每个候选描述词,依据公式$\text{DC}_j=\alpha*\text{Score(Keyword-DCF}^j)+\beta*\text{Score(Keyword-DCL}^j)$计算综合权重;

Step5.2: 将每个类簇$\text{C}_i(i\in[1,K])$中得分最高的候选描述词DC_i作为该类簇的聚类描述。

图2 基于DCF-DCL组合策略的聚类描述算法

该方法的基本思想为:首先利用DCF方法生成聚类描述词候选集合,再利用DCL方法生成聚类描述词集合,然后对两个集合进行基于相似度方法的类簇描述词加权归并操作,将每个类簇得分最高的候选描述词作为该类簇的聚类描述词。

图2给出基于DCF-DCL组合策略的聚类描述算法的详细描述。基于DCF-DCL组合策略的聚类描述算法主要包括DCF聚类描述预处理(高频关键词提取)、文本聚类、DCF聚类描述生成、DCL聚类描述生成、DCF-DCL聚类描述生成等5个步骤。

3 实验结果分析与讨论

本节依据图2所示的基于DCF-DCL组合策略的聚类描述算法,分别得到基于DCF、DCL1、DCL2、DCF-DCL1以及DCF-DCL2等五种聚类描述算法生成的聚类描述结果。

3.1 实验数据

以人大报刊复印资料[20]"人大2005年一季度经济类专题"库中的经济类论文2000篇作为数据集。数据集中的论文包括题名、摘要、关键词、带有段落和章节、图表标题信息以及参考文献等部分。对样本集合聚类后的类簇进行人工标注,将生成类别标注与对应类簇作为聚类描述算法的训练集和测试集。数据集主要包括类簇描述和对应类簇,每个类簇中包含属于该类簇的文本,而文本为学术论文形式,包括论文题名、摘要、段落、章节标题以及参考文献等信息。实验测评部分,本文采用10折交叉验证方法[10]。

3.2 评价方法

表1 聚类描述结果评价列联表

	人工标引为描述词	人工标引为非描述词
系统标引为描述词	a	b
系统标引为非描述词	c	d

聚类描述的结果可以表示如表1所示。本实验将人工标引的结果分为两种情况,即:人工标引为描述词的情况与人工标引为非描述词的情况,其中人工标引为非描述词的情况,就是将人工标引描述词后,类簇描述候选词语剩下的词作为非描述词。本文主要利用信息检索领域经典的评价方法,即查准率(P)、召回率(R)以及F1值对标引模型的标引性能进行评价,指标计算方法分别如下所。.

$$P=\frac{a}{a+b} \quad (1)$$

$$R=\frac{a}{a+c} \quad (2)$$

$$F1(P,R)=\frac{2PR}{P+R} \quad (3)$$

3.3 实验结果与分析

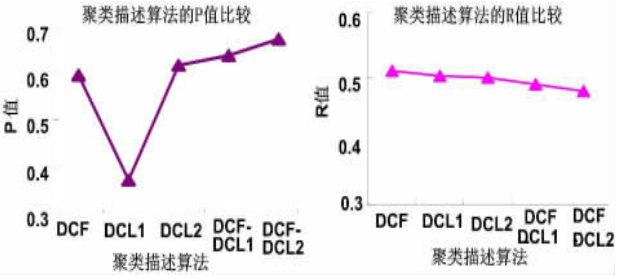
我们利用 3.1 中的所使用的数据集对这五种算法进行了 10 折交叉验证。

表 2 给出了 DCF、DCL1、DCL2、DCF-DCL1 以及 DCF-DCL2 等五种聚类描述算法的聚类描述结果的 P 值、R 值、F1 值结果。图 3(a)、(b)、(c)分别为五种聚类描述算法描述结果的 P、R、F1 值比较图。

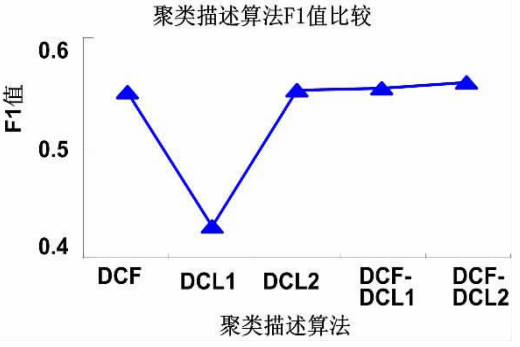
由图 3(a)可知，从聚类描述的查准率角度来看，基于 DCF-DCL 组合策略的聚类描述算法都要优于不组合的情况，其中 DCF-DCL2 组合策略优于 DCF-DCL1 组合策略。五种聚类描述算法中，传统的中心向量表示法，即 DCL1 的查准率最低，其次为 DCF 方法。由图 3(b)可知，从聚类描述的召回率角度来看，DCF 方法是最优的，其余依次为 DCFL1、DCL2、DCF-DCL1、DCF-DCL2。但它们之间的差异不是非常的明显。由图 3(c)可以看出，综合查准率和召回率后，基于 DCF-DCL 组合策略的稍优于不组合的情况，其中 DCF-DCL2 组合策略优于 DCF-DCL1 组合策略，其余依次为 DCL2、DCF、DCL1。传统的中心向量方法的描述性能均要次于其他描述算法。另外，通过图 3 (c) 还可以看出，DCF、DCL2、DCF-DCL1、DCF-DCL2 这四个聚类算法的 F1 值差异不是很显著，但结合查准率来看，可以看出它们之间的明显差异。聚类描述任务中的聚类描述要求具有一定精确度和区分性，即聚类描述任务主要偏向于获得每个类簇的精确描述，因此我们认为在聚类任务中，基于 DCF-DCL 组合策略的聚类描述算法要优于不组合的情况。另一方面，基于 DCF-DCL 组合策略的聚类描述算法可以从一定程度上解决前面所指出的 DCF 聚类描述算法中存在的语义间隔与聚类描述直观解释问题。

表 2 五种聚类描述算法描述结果的 P、R、F1 值

算法	P	R	F1
DCF	0.59723	0.50894	0.54956
DCL1	0.37202	0.50276	0.42762
DCL2	0.61973	0.49877	0.55271
DCF-DCL1	0.63826	0.48982	0.55427
DCF-DCL2	0.67372	0.47683	0.55843



(a) P 值比较 (b) R 值比较



(c) F_1 值比较

图 3 五种聚类描述算法的 P 值(a)、R 值(b)、F_1 值(c)比较图

4 DCF-DCL 聚类描述方法在搜索结果聚类中的应用

搜索结果聚类是指对查询返回的结果进行实时聚类[13-14,17,21]。针对搜索结果这一特殊任务的聚类算法时间复杂度应该尽量低，但同时又必须保证聚类的质量，即要求类簇内部的文档的相关性，允许类簇间存在重叠，并且聚类描述具有较强的可理解性。

算法：基于主题的搜索结果聚类算法.

输入：查询式 Query，信息检索系统 IR-SYSTEM；

文档聚类参数：文档聚类的类簇数目 K；

聚类描述参数：DCF 与 DCL 聚类描述权重系数分别为 α、β，且 α+β=1；

输出：K 个带有类簇描述 DC_i 的类簇 C_i，$i\in[1,K]$；

步骤：

//Setp1： 返回的查询结果

Step1.1：将 Query 提交给 IR-SYSTEM，返回前 N 条最相关的文档，组合文档集合 $D=\{\vec{d}_1,\cdots,\vec{d}_j,\cdots,\vec{d}_n\}$

Step1.2：抽取每篇文档 $\vec{d}_j$ 的标题(title)与文摘(Snippet)；

//Step2： 文本聚类

Step2.1：利用 Fuzzy C-Means 等聚类算法对文档集合进行聚类，生成 D 的一个划分为 $C=\{C|\ C\subseteq D\}$，$D=\cup C_i$，$i\in[1,K]$，得到每个类簇的中心向量 CV_i，$i\in[1,K]$；

//Step3： DCL 聚类描述生成

Step3.1：利用 DCL1 或 DCL2 模型生成每个类簇 C_i 的聚类描述 DC_i，$i\in[1,K]$；

图 4 基于 DCF-DCL 聚类描述方法的搜索结果聚类算法

基于此，本文提出基于主题的的搜索结果聚类方法，即返回文档集合经过主题提取、Fuzzy C-Means 聚类、DCL 聚类描述算法后，得到最终的返回结果的聚类导航列表，如图 5(d)所示(标记为 NJU-TCE)。图 4 给出了基于主题的搜索结果聚类的算法描述。

(a) Vivisimo[22]　(b) Carrot2[23]

(c) BBMao[24]　(d) NJU-TCE

图 5　四种搜索结果聚类系统结果示例

图 5 给出了查询式为“知识组织”(搜索 BBMao 和 NJU-TCE)或“Knowledge Organization”(搜索 Vivisimo 和 Carrot2)时的搜索查询结果聚类的结果。

5　结　语

本文提出的基于 DCF-DCL 组合策略的聚类描述算法实质上是一种基于集成学习方法的聚类描述算法。本文采用最简单的投票学习的方法。由于基于 DCF-DCL 组合策略的聚类描述算法是依赖于 DCF 和 DCL 两种聚类描述算法的描述结果。因此，要提高聚类描述质量，必须从根本上分别提高 DCF 和 DCL 的聚类描述质量。而这两种聚类描述质量又依赖于文本聚类描述算法的质量、聚类描述的特征选择、聚类描述模型的优化等因素。对这些方面进行研究是我们今后进一步的工作。此外，对搜索结果聚类进行有效评估也是本文下一步要进行的工作。

参考文献

1 Tseng Y-H, Lin C-J, Chen H H, Lin Y-H. Toward Generic Title Generation for Clustered Documents [C]. In: Proceedings of the 3rd Asia Information Retrieval Symposium (AIRS2006), Singapore, 2006: 145-157.

2 Popescul A, Ungar L. Automatic Labeling of Document Clusters [R]. Unpublished manuscript, available at http://www.cis.upenn.edu/~popescul/Publications/popescul00labeling.pdf. Accessed, 2007-01-10.

3 Pucktada T, Jamie C. Automatically Labeling Hierarchical Clusters[C]. In: Proceedings of the 2006 International Conference on Digital government research, San Diego, CA, USA, 2006: 167-176.

4 Maqbool O, Babri H A. Interpreting Clustering Results through Cluster Labeling [C]. In: Proceedings of the IEEE International Conference on Emerging Technologies (ICET'05), Islamabad, Pakistan, 2005: 429-434.

5 Stein B, Meyer zu Eissen S. Topic Identification: Framework and Application[C]. In: Proceedings of the 4th International Conference on Knowledge Management (I-KNOW 04), Graz, Austria, 2004: 353-360.

6 Lawrie D, Croft W B, Rosenberg A L. Finding Topic Words for Hierarchical Summarization [C]. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'01), New Orleans, LA, USA, 2001: 249-357.

7 Muscat R. Automatic Document Clustering Using Topic Analysis[R]. Technical Report CSAI2005-01, Department of Computer Science & AI, University of Malta, 2005: 1-16.

8 Li H, Shen D, Zhang B Y, Chen Z, Yang Q. Adding Semantics to Email Clustering[C]. In: Proceedings of the IEEE 6th International Conference on Data Mining (ICDM 06), Hong Kong, China, 2006: 18-22.

9 Dawid W. Descriptive Clustering as a Method for Exploring Text Collections [D]. PhD Thesis. Poznan University of Technology, Poznań, Poland, 2006: 7-56.

10 Han J, Kamber M. Data Mining: Concepts and Techniques [M]. San Francisco: Morgan Kaufmann, 2001: 322-324, 376-379.

11 Glenisson P, GlÃ Ã nzel W, Janssens F, De Moor B. Combining Full Text and Bibliometric Information in Mapping

Scientific Disciplines [J]. Information Processing & Management, 2005, 41(6): 1548-1572.
12 Lai K K, Wu S J. Using the Patent Co-citation Approach to Establish a New Patent Classification System [J]. Information Processing & Management, 2005, 41(2): 313-330.
13 Cutting D R, Karger D R, Pedersen J O, Tukey J W. Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections [C]. In: Proceedings of the 15th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'92), Copenhagen, Denmark, 1992: 318-329.
14 Cutting D R, Karger D R, Pedersen J O. Constant Interaction-time Scatter/Gather Browsing of Large Document Collections [C]. In: Proceedings of the 16th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'93), Pittsburgh, PN, USA, 1993: 126-135.
15 Muller A, Dorre J, Gerstl P, Seiffert R. The TaxGen Framework: Automating the Generation of Taxonomy for a Large Document Collection [C]. In: Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS1999), Maui, HI, USA, 1999: 2034-2042.
16 Anton V L, Croft W B. An Evaluation of Techniques for Clustering Search Results [R]. Technical Report IR-76, Department of Computer Science, University of Massachusetts, Amherst, 1996: 1-19.
17 Zamir O, Etzioni O. Web Document Clustering: A Feasibility Demonstration[C]. In: Proceedings of the 19th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'98), Melbourne, Australia, 1998: 46-54.
18 Glover E, Pennock D M, Lawrence S, Krovetz R. Inferring Hierarchical Descriptions [C]. In: Proceedings of the 11th International Conference on Information and Knowledge Management (CKIM2002), McLean, VA, 2002: 4-9.
19 章成志. 主题聚类及其应用研究 [D]. 南京: 南京大学, 2007: 136-143.
20 人大报刊复印资料 [EB/OL]. http://www.zlzx.org, 2007-12-01.
21 Zeng H J, He Q, Chen Z, Ma W Y, Ma J. Learning to Cluster Web Search Results [C]. Proceedings of 27th Annual International Conference on Research and Development in Information Retrieval, Sheffield, United Kingdom, 2004: 210-217.
22 Vivisimo Clustering Engine [EB/OL]. http://vivisimo.com, 2006-05-01.
23 Carrot Clustering Engine[EB/OL]. http://demo.carrot2.org, 2006-05-01.
24 BBMao 社会化搜索引擎 [EB/OL]. http://www.bbmao.com, 2006-05-01.

(责任编辑:徐 波)

(上接第 1061 页)

流媒体技术作为一项新的网络技术，一旦与原有的网络技术结合起来，将给我们的网络应用带来新的发展和变化。我们完全可以相信,流媒体技术的应用会使校园网变得更加精彩，使数字图书馆的工作更加完善。

参考文献

1 张 程，朱庆生. 采用流媒体技术实现网络中的视频和音频传播[J]. 计算机工程与设计, 2002, (2): 57-59.
2 黄永跃. 流媒体技术在现代图书馆中的应用[J]. 现代图书情报技术, 2003, (1): 76-78.
3 张 丽. 流媒体技术大全 [M]. 北京: 中国青年出版社, 2001: 101-105.
4 胡 俊. 流媒体技术在数字图书馆中的应用 [J]. 情报科学, 2001, (4): 351-354.
5 吴雄林，李 勇. 流媒体技术在图书馆声像信息服务中的应用[J]. 图书情报知识, 2003, (4): 51-53.
6 赵彦龙. 流媒体技术在图书馆数字化服务和建设中的应用[J]. 图书馆工作与研究, 2004, (2): 15-16.
7 厉 励, 张宏坡, 李 海, 周 兵. 基于 QOS 的磁盘调度策略[J]. 计算机科学, 2006, (9): 118-119.
8 孙 为, 张宝杰, 车 嵘. 基于大数据量实时流媒体 P2P 树算法研究[J]. 兰州理工大学学报, 2006, (6): 109-110.
9 杨飞飞. 流媒体技术在图书馆多媒体资源数字化中的应用[J]. 图书馆, 2007, (6): 78-80.
10 高先锋, 张洪沼. 图书馆海量存储系统架构与接口的选择[J]. 现代图书情报技术, 2003, (4): 81-83.

(责任编辑:徐 波)