●梁 娜（中国科学院文献情报中心 北京 100080；中国科学院研究生院 北京 100049）

知识组织体系登记系统

摘 要：知识组织体系登记系统是描述、揭示、组织知识组织体系的重要工具。本文通过对现有的几种知识组织体系范例登记系统的分析，把握知识组织体系登记系统领域的发展现状，研究了 ISO/IEC 19763 对知识组织体系登记系统的解决方案，最后分析了知识组织体系登记系统需要重点解决的问题。

关键词：知识组织体系；本体；登记系统

Abstract: Knowledge organization system registry is an important tool to describe, identify and organize knowledge organization system. The paper analyzes several typical knowledge organization system registries to grasp the development status of this field, and studies the solutions of ISO/IEC 19763 to knowledge organization system registry. Finally, several important issues that should be dealt with about knowledge organization system registry are analyzed.

Keywords: knowledge organization system; ontology; registry

知识组织体系描述和组织了知识概念和知识概念间的相互关系。在分布式信息环境中存在多种类型、不同应用领域的知识组织体系。它们之间可能包含相同或相近的知识概念、知识内容，存在关联、相似、继承等逻辑关系，同样知识概念和知识内容间也存在类似的关系联系。要利用知识组织体系及其所包含的内容概念和关联关系，就要对这些对象的定义、结构、应用方式（解析、复用、继承、转换、集成等）、关系等进行描述、发布，让应用系统能够对知识组织体系进行搜寻、定位、析取、交换和集成，从而利用知识组织体系挖掘知识内容、组织开放信息系统和开放信息服务。知识组织体系登记系统以特定的数据模型组织了知识组织体系及其知识内容和相互关系的描述信息和利用信息，并提供一系列功能实现对以上对象和对象信息的检索、发布、调用和集成。

1 知识组织体系登记系统范例分析

1.1 XMDR

XMDR（eXtensible Metadata Registry）[1]扩展元数据登记系统，是美国劳伦斯伯克利国家实验室（Lawrence Berkley National Laboratory）新近启动的研究项目，该项目支持对元数据登记系统（Metadata Registry）中数据元素、术语、概念结构的语义的利用（例如存储、检索和呈现等）。XMDR认为现有的元数据登记系统标准（如 ISO/IEC 11179：Information Technology —Metadata Registries）[2]对知识组织体系（术语集、分类表和本体等）不重视，XMDR 希望提供一个针对知识组织体系的可描述的逻辑表示形式来获取知识组织体系中的语义定义。

XMDR 对所登记的元数据的内容进行了扩展[3]，不再仅仅限于描述数据元素的元数据，还包括描述和组织内容概念和相互关系的知识组织体系，例如分类体系、术语集、编码集、本体等。对描述数据元素的元数据而言，这样的登记系统比较容易实现，而通过登记系统去揭示和展现后者中存在的关联、相似、集成等语义逻辑关系是比较复杂的，于是 XMDR 采取的办法就是用图形/图表来表示。对于分类体系、术语集、编码集这类有着层级聚合和分类体系的知识组织体系，它采用树（Tree）、分面图（Faceted Classification）、有序树（Ordered Trees）来表示其中的等级、分面分类、继承等关系；对于本体，它采用直向循环图（Directed Acyclic Graph）、格子（Lattices）、双向图（Bipartite Graphs）等来展现本体中的对象类的层级体系、对象类之间的逻辑相互关系，以及按照相互关系来表现对象类的属性和属性取值限制。

XMDR 这种用图形理论来描述知识组织体系的方法的好处在于，由于图形/图表所展现的元素和元素间关系的完整性和限制性较强，它易于理解元数据结构，可以与相应的元数据集绑定，并且有支持查询和处理图形/图表的运算法则。但其中的难点很多，例如如何登记和管理多种图形、图表结构；如何查询图表结构，采用什么标准的查询语言；在图形/图表比较复杂的情况下如何执行查询结果；如何建立图形/图表间的关联等。目前该项目还处于初期阶段，没有实际的系统可表现，但提供了一种描述和组织知识组织体系的全新思路。

1.2 AOS Ontology Registry

联合国粮农组织农业本体服务项目（*UN FAO Agricul-*

tural Ontology Service (AOS) Project)[4]，可看作是农业领域各种知识组织体系的联合存储系统和组织系统，它以粮农组织的多语种农业叙词表（AGROVOC）为基础，集合了多个权威农业机构的词表，如美国国家农业图书馆的AgNIC叙词表、国际农业和生物中心的CAB叙词表等。

AOS包括两个部分：AGROVOC概念服务器（AGROVOC Concept Server）和本体登记系统（Ontology Registry）。前者包含了农业领域各个子学科的所有术语、概念和相互关系，提供了一种类似于“积木”的基础材料，利用它可以构建描述农业信息资源的元数据本体和农业领域的学科子集内的领域本体，例如林学、渔牧学等，再引入本体登记系统。AOS本体登记系统的作用就是存储这些领域本体，目前已有一个渔学本体和三个粮食安全本体。但目前这样的本体登记系统过于简单，只是一个存储库，没有对本体进行描述，也没有登记系统的相关功能。但从管理的角度来看，AOS Ontology Registry可以作为一个分布式农业本体的管理系统，粮农组织中任何农业机构如果想创建本体时，可以利用这个登记系统来查找相关可利用的本体，避免不同的机构可能会针对同一主题领域创建不同的农业本体，从而更能促进农业科学领域内容概念的共享和信息系统的交互理解。

1.3 FZI Ontology Registry

FZI Ontology Registry[5]是德国Karlsruhe大学FZI研究中心分布式本体复用研究中的本体登记系统，该登记系统支持e-Business环境下不同本体的查找和复用。该登记系统的内容就是关于本体的描述信息，包括本体实例模型概念(OIMODEL Concept)，描述了本体的创建时间、本体名称等；个人与组织概念（PERSON or ORGANIZATION Concept)，描述了本体的创建者和机构；本体相关项概念(TERM Concept)，描述了本体与特定项目、特定应用领域的联系等信息。

FZI Ontology Registry支持的本体查找方法有3种：描述信息的匹配检索，主要根据本体的描述信息进行检索；Query-By-Example（QBE）实例查询，例如提供本体的领域限制信息来限定本体实例的查找范围；本体中包含的相关概念项的关键词定义。对于第3种方法，由于不同的本体可能包括相同的概念和内容，所以它使用Wordnet[6]这种最常用的英语词汇叙词表来帮助优化检索。具体方法是首先检索出与检索项匹配的所有本体，再将每个检索项与Wordnet中的词汇匹配，检索出一个包含所有可能相关语义的词汇列表，同时匹配项与检索项的差异也被列出，结合本体的描述信息与限制信息，将过滤出最终匹配的本体。

而对Ontology的复用，FZI Ontology Registry主要是利用以上方法，检索Ontology Registry中支持构建所需的本体的定义信息，通过登记系统的指向信息可以直接从保存实际本体的服务器上复制可用的本体，从而复用本体中的可用的概念和定义。

1.4 UMLS Metathesaurus

UMLS元叙词表[7]（UMLS Metathesaurus）是美国国家医学图书馆的统一医学语言系统（Unified Medical Language System，UMLS）的核心部分，收录了100多个医学领域受控词汇表和分类表的超过100万个生物医学概念和500万个概念名称。UMLS元叙词表规定了这些概念的语义类型和关系，并通过复合方法把各种各样的概念的属性联系起来。UMLS语义网络[8]（UMLS Semantic Network）定义了概念类别和类别间的语义关系，为UMLS元叙词表中的概念提供了固定的语义类型。

UMLS元叙词表提供了概念的3类信息：名称信息，包括概念唯一标识符、概念名称、术语状态、语种等；属性信息，包括概念定义、语义类型、表达概念的术语和词串、来源词表及词类；关系信息，包括相关概念关系（上位、下位、相似等关系）、组配表达信息（与概念相关的多术语组配表达式）、共现概念信息（来自同一信息源中共同出现的概念及共现频率统计数据）。UMLS元叙词表提供3种索引支持对概念名称信息的检索，包括：字索引(Word Index)，它的一个条目对应一个表达概念的词串，条目中包含了概念唯一标识符、术语唯一标识符和词串唯一标识符；规范化字索引（Normalized Word Index）和规范化词索引（Normalized String Index)。通过这3个索引，可以将UMLS元叙词表中词串和与之相关的术语标识符和概念标识符链接起来，从而把表达同一概念的词语联系起来。

从以上可以看出，UMLS元叙词表在一定意义上实现了知识组织体系登记系统的组织和描述知识组织体系的作用，并以索引的方式实现对词表中概念和相互关系的查找，支持了特定关系的概念间的联系。

2 描述知识组织体系登记系统的方案和标准规范

2.1 NKOS Registry

网络知识组织体系工作组（Networked Knowledge Organization Systems (NKOS) Workshop）2001年发布了NKOS登记系统数据元素草案V 3.0[9]，在沿用DC的15个基本元素的基础上，提供了描述知识组织体系的元数据框架。以ISO/IEC 11179规定的10个属性作为描述数据元素的标准，定义了20个描述元素，包括KOS题名（KOS Title)、可选题名（Alternative Title)、创建者（Creator)、KOS主题(KOS Subject)、描述（Description)、发布者（Publisher)、

时间（Date）、KOS 类型（KOS Type）、格式（Format）、标识符（Identifier）、语言（Language）、KOS 关系（KOS Relation）、权限（Rights）、实体类型（Entity Type）、实体取值（Entity Value）、关联关系（Relationships）、排列（Arrangement）、应用（Application）和次要主题（Minor Subject）。需要强调其中 3 个，KOS Type 指出了知识组织体系本身的内容类型，可以是规范档、分类表、叙词表等；KOS Relation 指明了与知识组织体系关联的参照体系，例如 URI、DOI 数字对象标识系统、ISBN 国际标准书号体系，而指明知识组织体系中相互关系的是 KOS Relationship 元素，它给出了知识组织体系中实体间的继承、相联、相似等关系。

2.2 ISO/IEC 19763-3：Metamodel Framework for Ontology Registry

元模型互操作框架标准（ISO/IEC 19763：Information Technology—Framework for MetaModel Interoperability）[10] 拟解决应用系统间互操作问题，见图 1。两个应用系统间在不知道对方是谁，能提供何种服务及其具体语义的情况下，是难以交互理解的。可以建立各自应用系统的本体，通过本体对自我系统进行描述和揭示，并将本体的描述信息在公共的本体登记系统中发布供查找和发现，从而实现基于本体的应用系统交互。由于本体登记系统只组织本体的描述信息，而实际的本体则存储在本体存储库（Ontology Repository）中，所以这样的交互还依赖于本体登记系统与本体存储库之间的互操作。

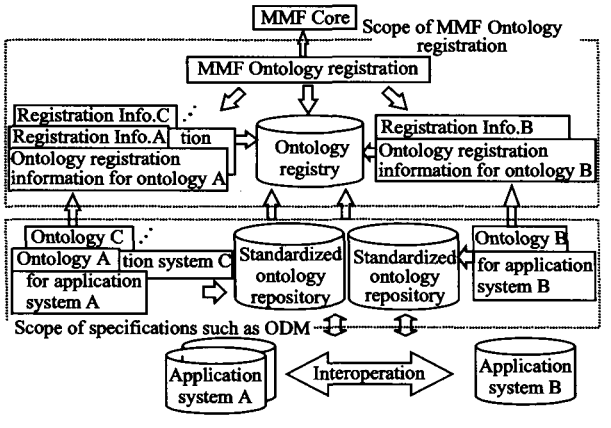

图 1　元模型互操作框架

该标准的第三部分（ISO/IEC 19763-3：Metamodel for Ontology Registration，在该标准中简称为 MMF Ontology Registration）[11]，它的主要任务就是提供本体登记系统的元模型。由于不同的应用系统的本体是不同的，而且不同的本体可有不同的表示描述语言和方式。在这种情况下，MMF Ontology Registration 为促成本体间的理解，提供了两种类型的本体，包括参照本体（Reference Ontology）和本地本体（Local Ontology）。前者是关于某主题领域的已预定义的标准本体，后者是基于前者构建的应用系统的本体。在参照本体的指向和帮助下，不同的本地本体之间可以交互理解。同时 MMF Ontology Registration 明确了 Ontology 的基本结构包括 3 个部分，分别是本体（Ontology）、本体组分（Ontology Component）和本体原子结构（Ontology Atomic Construction）。这 3 个部分分别指向实际本体描述语言中的本体（Ontology）、语句（Sentence）和符号（Symbol），见图 2。那么 Ontology 的不同表示方式、描述语言（如 OWL、TM、XCL、UML、RDF 等）中都可找到这 3 个基本部分的描述信息。

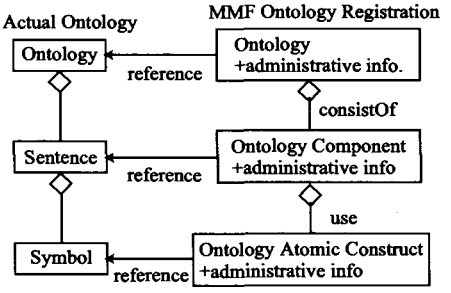

图 2　元模型框架本体登记系统结构

对本体的描述，MMF Ontology Registration 建立了一个描述元模型（见图 3），包括 15 个类，有领域本体（onto-domain）、本体实例（onto-instance）、本体选择条件（onto-selection）、本体概念（onto-concept）、本体结构（onto-construct）、本体类型符（onto-classifier）、来源本体（source-onto）、本体原子结构（atomic-onto-construct）、来源本体组分（SO-component）、本体演变信息（evolution-info）、OWL 语言表示（OWL-expression）、TM 语言表示（TM-expression）、RDFS 语言表示（RDFS-expression）、来源本体组分变形（SOC-variant）和本地本体（local-onto）。在这个元模型中，一个本体实例产生于特定的领域本体，从本体实例中选取特定条件和关系可以表达本体中的概念，而这些概念与领域本体紧密相关。领域本体有它的本体来源和表现模型，可以是 RDF、OWL、TM。从本体实例中还可抽取出局部本体。本体来源的组成部分可生成多种本体模型，支持本体实例的构建。

MMF Ontology Registration 提供了一个接口（Interface）来实现本体登记系统与实际的本体存储库之间进行交互，这个接口由本体定义元模型（Ontology Definition Metamodel，ODM）[12] 来实现。ODM 是本体的模型化表现，就是对本体抽象简化，转换为模型。ODM 包含多种 Ontology 表示模型（如 OWL、RDF、UML、TM、ER），是一个本体的元模型体系，包括 6 个部分（见图 4）。这些元模型都不是支持谓词表示的完全揭示性谓词演算语言，所以 ODM 又引入了简易通用逻辑（Simple Common Logic，SCL）。ODM 中各种元

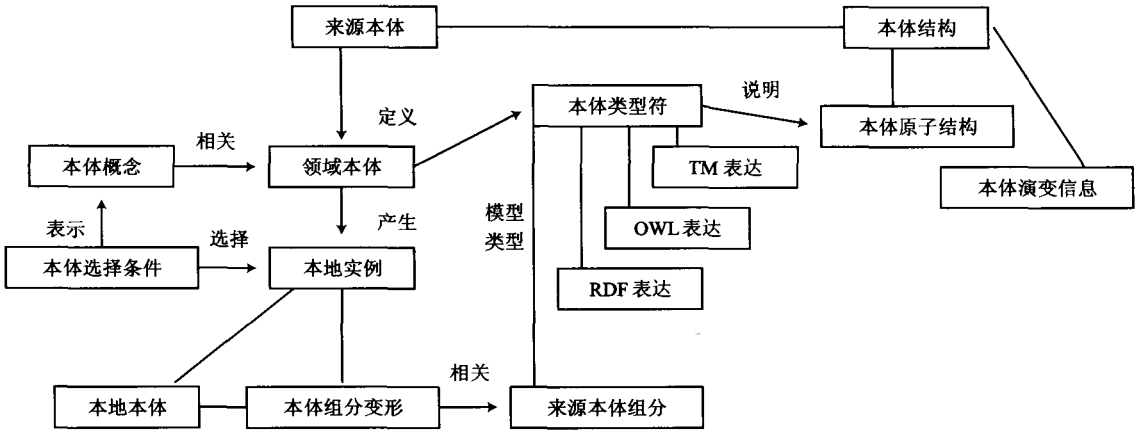

图 3 本体登记系统元模型

模型的连接通过 OMG QVT（Query | View | Transform）系统的映射表示来实现。QVT 可以实现对模型的查询、查看和转换。为避免 n 次映射，ODM 还建立了描述逻辑（Description Logic，DL）作为元模型间映射的双向映射，如 UML 与 OWL 之间的映射，须先建立 UML 与 DL 的映射，再从 DL 映射到 OWL。SCL 比其他元模型更具有描述性，可以表示 Ontology，但难以与其他元模型建立映射，所以 ODM 将 SCL 用于实现谓词，这就是说，谓词可以在一个“Primary Metamodel”中定义描述，如 OWL，但是在 SCL 中执行实现。而在这个“Primary Metamodel”中定义的其他相关元素也能映射到 SCL 中，所以 DL 与 SCL 间的单向映射也能建立起来。

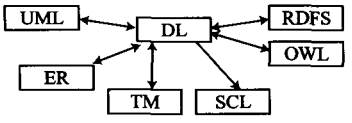

图 4 ODM 的基本结构

3 知识组织体系登记系统的功能需求分析

上述知识组织体系登记范例系统能够立足某主题领域，组织基于叙词表的领域本体的构建实践，探索利用登记系统查找和复用知识本体的方法。但都是从系统的角度看待知识组织体系登记系统的作用，没有从描述的角度揭示知识组织体系本身。NKOS 的元数据草案和 ISO/IEC 19763-3 则提供了较为规范的描述知识组织体系和本体的核心元素和元模型。但前者仅是一个知识组织体系本身的描述框架，并未充分揭示知识组织体系中概念间的相互关系。后者揭示了本体间的一些关系（如来源本体、领域本体、本体实例、本体组分和本体原子结构间的关系）、本体的描述方法（如 RDF，OWL，TM）、本体的变形/变化信息、本体变形后与来源本体和本体组分间的关系，以及本体登记系统的接口利用信息（如 ODM）等。但这样的标准规范仅限于本体的描述。

所以知识组织体系登记系统应包括知识组织体系的描述信息、知识组织体系间关系、知识组织体系的描述方法、知识组织体系的注释及版本信息、知识组织体系的变化信息、知识组织体系间的差异信息、知识组织体系系统和接口信息、知识组织体系开发和利用工具信息以及知识组织体系的应用环境信息等。在功能上，登记系统要支持以上信息的浏览、检索、发布；支持知识组织体系的匹配与查询；支撑知识组织体系的调用、集成。同时知识组织体系登记系统的功能框架还应包括知识内容间关系的形式化表现功能，例如可视化技术以具体的显性的方法来支持知识内容的挖掘、发现，以及知识组织体系及其实例的检索等。

4 小结

知识组织体系登记系统不仅仅是一个存储和管理知识组织体系的工具，更重要的是，它提供了一种发布、查询、发现知识组织体系及其内容概念的服务，利用这种服务可以发现知识组织体系中内容概念及其相互关系，利用这些关系可以构建、复用知识组织体系。但目前知识组织体系登记系统要解决的问题包括：①现有的登记系统未能充分揭示知识组织体系中概念间的关系；②知识组织体系是动态的、变化的，利用登记系统如何记录、跟踪和监测知识组织体系的变化；③缺少方法和工具来支持和实施知识组织体系的匹配、复用；④知识组织体系登记系统与知识组织体系存储库的显著区别是什么；⑤具体以何种形式

（下转第 460 页）

根据资源描述文件，利用余弦相似性计算资源的属性相似性。

$$I_{attribute}(p, q) = \frac{\sum_{k=1}^{d} S_{p,k} \times S_{q,k}}{\sqrt{\sum_{k=1}^{d} (S_{p,k})^2 \times \sum_{k=1}^{d} (S_{q,k})^2}}$$

将基于用户评价矩阵的资源相似性和基于资源描述文件的属性相似性用线性组合表示为：

$$s(p, q) = aI_{evaluation}(p, q) + (1-a) I_{attribute}(p, q)$$

则 s（p，q）为资源 p 和 q 的相似性，a 为权重系数，如果 a=0，则 s（p，q）$= I_{attribute}$（p，q），即相似性仅从资源描述文件获得；如果 a=1 时，相似性就会基于用户评价矩阵的相似性，s（p，q）$= I_{evaluation}$（p，q），则同样的预测评价值为：

$$p_{a,k} = \frac{\sum_{t \in I_k} [r_{a,t} \cdot s(k, t)]}{\sum_{t \in I_k} s(k, t)}$$

需要注意的是，这里所命名的用户—资源相似性并不是计算用户同资源的相似性，而仍是资源相似性，所不同的是利用了资源描述文件。

4 结论

本文所提出的 3 种个性化推荐策略，在数字图书馆领域都可以得到相应利用：对于频繁使用数字图书馆的用户，可以视为目标用户，利用用户—用户相似性推荐算法；对于非频繁使用数字图书馆的用户，可以利用资源—资源相似性推荐算法，向其推荐目标资源；而为了提高个性化推荐的精度，我们可以采取用户—资源相似性推荐算法。对于这 3 种推荐策略，数字图书馆在具体应用时要综合进行考虑和选择。

还需要指出的是，数字图书馆的个性化推荐还面临着很多需要改善的问题，即对用户描述文件的不断完善，对资源描述文件中本体库的建立，以及协同过滤技术带来的稀疏性、冷开始问题，这些问题的妥善解决都能够相应地提高个性化推荐的准确度，从而真正实现数字图书馆中“以用户为中心”的个性化推荐服务。□

参考文献

1 廖亚莉，王锡钢，战学刚．电子商务的个性化服务．鞍山科技大学学报，2004，27（3）：194～197，201

2 余力，刘鲁，罗掌华．我国电子商务推荐策略的比较分析．系统工程理论与实践，2004（8）：96～101

3 Schafer J，Konstan J，Riedl J．E-commerce Recommendation Applications．Data Mining and Knowledge Discovery，2001（5）

4 Breese J S，Heckerman D，Kadie C．Empirical Analysis of Predictive Algorithms for Collaborative Filtering．In：Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence，1998．43～52

5 Goldberg D，et al．Using Collaborative Filtering to Weave an Information Tapestry．Communications of the ACM，1992，35（12）

6 刘柏崇．基于本体的数字图书馆信息过滤研究．上海交通大学学报，2003，（37）：171～175，183

7 余正涛，宋丽哲，樊孝忠．基于本体的个性化领域信息服务．计算机工程，2005，31（5）：22～24，81

8 Sarwar B，et al．Item-based Collaborative Filtering Recommendation Algorithms．In：Proceedings of the Tenth International World Wide Web Conference，2001．285～295

作者简介：李君君，女，1980 年生，博士生。发表论文 5 篇，出版著作 1 部。

叶凤云，女，1980 年生，硕士生。发表论文 1 篇。

收稿日期：2006－02－20

（上接第 471 页）

和方法来存储和组织知识组织体系本身及其中内容概念，等等。□

参考文献

1 XMDR（eXtensible Metadata Registry）．http：//xmdr．org/，2005-01-15

2 ISO/IEC 11179：Information Technology—Metadata registries．http：//metadata-standards．org/11179/，2005-01-15

3 XMDR Overview．http：//hpcrd．lbl．gov/SDM/XMDR/overview．html，2005-01-15

4 Agricultural Ontology Service（AOS）．http：//www．fao．org/agris/aos/，2005-01-15

5 Maedche A，et al．An Infrastructure for Searching，Reusing and Evolving Distributed Ontologies．http：//www2003．org/cdrom/papers/refereed/p104/p104-maedche．html，2005-01-15

6 Wordnet．http：//www．cogsci．princeton．edu/～wn/，2005-01-15

7 UMLS Metathesaurus．http：//www．nlm．nih．gov/pubs/factsheets/umlsmeta．html，2005-01-15

8 UMLS Semantic Network．http：//www．nlm．nih．gov/pubs/factsheets/umlssemn．html，2005-01-15

9 Networked Knowledge Organization Systems（NKOS）Registry Draft Proposal for Data Elements Version 3．http：//staff．oclc．org/～vizine/NKOS/Thesaurus_Registry_version3_rev．htm，2005-01-15

10 ISO/IEC 19763：Information Technology—Framework for MetaModel Interoperability．http：//metadata-standards．org/19763/，2005-01-15

11 ISO/IEC 19763-3：Metamodel for Ontology Registration．http：//jtc1sc32．org/doc/N1301-1350/32N1308T-CD19763-3．pdf，2005-01-15

12 ODM（Ontology Definition Metamodel）．http：//codip．grci．com/odm/draft/submission_text/ODMPrelimSubAug04R1．pdf，2005-01-15

作者简介：梁娜，女，博士生。发表论文 10 余篇。

收稿日期：2006－03－02