



个性化服务中跨系统用户建模方法研究 综述*

牛亚真^{1,2} 祝忠明¹

¹(中国科学院国家科学图书馆兰州分馆 兰州 730000)

²(中国科学院研究生院 北京 100049)

【摘要】系统地分析和总结跨系统用户建模的主要方法:基于统一用户模型的方法,主要是试图通过构建标准的本体或者通用的用户模型来满足不同系统的个性化需要;基于用户模型间映射和融合的方法,主要通过一定的映射规则对不同系统中用户模型进行映射和融合,实现跨系统的个性化服务;分布式开放用户建模方法,主要是基于社交网络、互操作、关联开放数据等来构建用户模型。通过对这些方法的比较分析,指出跨系统用户建模的发展趋势。

【关键词】语义网 跨系统用户建模 用户模型 社交网络 分布式 互操作 关联数据 关联开放数据

【分类号】G250.76

Overview about the Methods of Cross – system User Modeling for Personalization Service

Niu Yazhen^{1,2} Zhu Zhongming¹

¹(The Lanzhou Branch of National Science Library, Chinese Academy of Sciences, Lanzhou 730000, China)

²(Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

【Abstract】 This paper summarizes and analyzes the main methods of cross – system user modeling. The first one is a top – down approach, involving standard Ontologies or unified user models; the second research direction is a bottom – up approach based on mappings between different user model representations; the third one is distributed and open user modeling, based on the Social Web, interoperability and LOD. Finally, it points out the tendency of the cross – system user modeling.

【Keywords】 Semantic Web Cross – system user modeling User model Social Web Distributed Interoperability Linked data Linked open data

1 引言

个性化服务的一般过程可以概括为:收集与用户相关的数据;根据收集到的数据推断出用户的特征喜好和需要;对系统进行定制,满足用户的需要。其中,用户建模是个性化服务的核心和基础。一般地,系统单独地构建各

收稿日期:2012-03-21

收修改稿日期:2012-05-07

* 本文系中国科学院西部之光联合学者项目“机构知识库的语义增强方法与技术研究”的研究成果之一。

自的用户模型,同一用户在不同系统中会具有不同的用户模型。这就造成:同一用户的信息(如登记注册信息等)会在不同系统中重复出现,造成数据冗余;当新用户登录系统时,由于系统中没有或者只有很少的与新用户相关的数据,容易导致冷启动(Cold Start)问题^[1]。构建跨系统的用户模型可以很好地解决上述问题。跨系统的用户建模是在多应用场景中,以用户为中心构建能够满足不同系统需要的用户模型^[2]。利用这种方法可以更有效地捕捉和组织用户的兴趣和爱好,同时实现不同系统中用户数据的共享和重用。

2 跨系统用户建模的主要方法

2.1 基于统一用户模型的方法

这是一种自上而下(Top-down)的方法,试图定义标准的本体或者通用的用户模型来满足不同系统的个性化需要,主要关注用户模型本身的可重用性。早在20世纪末,已经有很多学者对此进行了相关的研究^[3]。下面介绍和分析几个典型实例:

(1) OntobUM

OntobUM是一个基于本体的统一用户模型框架,由Razmerita等^[4]于2003年提出,试图以本体作为构建用户模型的标准。OntobUM框架集成了三个本体:用户本体(定义了用户的不同属性及其关系)、领域本体(定义了特定领域和应用的的概念及其关系)、日志本体(定义了用户与应用系统之间的交互产生的语义关系)。为了使用户模型通用,OntobUM把数据存储成RDF/RDFS的格式。该模型通过用户模型的本体编辑器(UPE)提供显性定义,生成用户本体;同时,通过智能服务(Intelligent Services)工具更新和维护用户模型产生的隐性数据,提供个性化服务。

这种方法主要关注用户模型本身的可复用性,目标在于开发出一种支持用户模型重用的描述语言和表示方式。该方法在知识管理系统(KMS)中得到了应用,主要是试图构建一种可以被不同系统接受的,从而使由本体集成的用户模型在不同系统中适用。OntobUM也可以应用于其他不同领域的本体环境中,这充分体现了本体的可重用性。但是,由于本体设计是依赖于用户的知识和研究者的经验,设计本体的精确性并不高。

(2) UUCM

Niederée等^[5]于2004年提出了UUCM(Unified Us-

er Context Model),它是一种基于本体的统一用户情境模型,不仅描述了用户的相关特征,而且结合了用户的工作情境信息。其中,对于关系的描述,既描述了用户之间的关系,还描述了给定领域中的资源之间的关系。该模型基于“情境护照(Context Passport)”来抽取和聚合不同的个性化系统收集的用户数据,通过跨系统交流协议CSCP(Cross-System Communication Protocol)交换用户信息,是一个可扩展的多维度用户模型。同时,还用到了“元模型(Meta-model)”,这实际上是一种可共享的本体,不同的系统正是依赖此共享本体来实现个性化的服务。

“情境护照”是一种隐喻性的说法或概念,它是把用户及其情境信息封装为“Context Passport”,使之可以伴随用户在跨系统之间流转和使用,通过出示“情境护照”,提供用户情境信息,从而获得情境化敏感的或个性化的服务。这种方法要求所有的个性化应用系统必须依照UUCM框架构建用户模型,并且支持CSCP协议,这在实行中有很大的限制。研究者们试图利用机器学习技术来解决这种限制,但是这需要大量的用户访问不同系统,产生有规则可循的用户群数据。

(3) GUMO

Heckmann等^[6]于2005年提出了GUMO(General User Modeling Ontology),作为在语义网环境下分散的用户模型的一种统一表示。该方法是采用OWL表示用户模型的词汇和相互关系,OWL本体语言为定义本体提供了一种富表达力的语言,能够捕捉领域知识的语义。GUMO中的用户模型是根据UserML^[7]设计的,UserML把用户模型的维度分解成Auxiliary、Predicate、Range三个部分。例如,某用户的兴趣是足球,可以表示成:auxiliary = hasInterest, predicate = football, range = low-medium-high。

利用GUMO的优势在于它的语义一致性很强,用户模型数据被融合入本体中,并且以可折叠树的形式浏览和访问。把本体引入到用户建模中,给用户模型带来了丰富的语义,极大地提高了系统的自适应性^[8]。GUMO利用OWL构建复杂的用户模型概念,并以图的层级形式表示出来,这样可以表达相当全面的用户模型的特征维度。目前,GUMO已经在用户模型服务和泛在服务中有了广泛的应用。

从上面三个实例可以看出,为了满足不同系统间

信息的交换,这些方法引入了本体的概念,利用本体的共享性和可重用性构建统一的用户模型。然而,所有的应用系统采用统一的用户模型,意味着用户特定领域的特征不能够被充分地表达。

2.2 基于用户模型间映射和融合的方法

这是一种自底向上(Bottom-up)的方法,通过映射技术对不同系统中用户模型的数据进行聚合,完善用户模型,为系统提供更丰富的用户信息,从而满足不同系统间的个性化服务。下面介绍和分析两个典型实例:

(1)GUC

GUC(Generic User-model Component)是一个利用语义网技术表达用户模型的通用元件,由 Sluijs 等^[9]在 2005 年提出。它既提供存储用户数据模型的功能,又支持模型中用户数据的交换。GUC 的每一部分都有明显的界线,应用程序和用户是相互独立的。对每一个应用程序,都有一个模式映射到联合的本体;对每个用户,都有一个 UAV 知识库(User Application-View Repository)存储用户的信息。该方法通过一个描述不同系统用户模型的数据结构的模式,使用匹配和融合技术构建一个可共享的用户模型;同时,通过基于数据调谐(Data Reconciliation)规则的模式映射来完成不同系统间的数据交换。

模式映射方法在 GUC 中是应用的关键,很好地起到了交换数据的作用。GUC 很容易配置,主要应用于自适应网络信息系统(AWIS),可以与 AWIS 关联在一起。实施 GUC 框架强调配置的灵活性,然而目前 GUC 的可扩展性并不突出。不同的应用系统必须经过订阅才能向 GUC 上传数据,这给很多系统造成了不便。另外,用户需要知道哪部分信息被应用,所以用户的隐私权和授权认可问题需要解决。

(2)SUM

2005 年,González 等^[10]提出了一个基于多代理的智能用户模型 SUM(Smart User Model),用来支持异构的分布式环境中的跨领域的个性化推荐。在该模型中,使用属性-值(Attribute-Value)对的集合来表示用户特征,利用智能代理技术获得单个的用户模型,然后对用户模型进行聚合。SUM 不仅从推荐系统中学习用户的特征,还可以把用户的特征传递给其他的推荐系统。为了使 SUM 可以在不同领域的系统中应用,首先定义应用领域 i 中的用户模型(UM_i)。通过加权图 $G(SUM_i, UM_i)$ 构建 SUM 和 UM 之间的关系,并且由加权图数据填充用户模型。

该模型中结合了智能自适应控制系统(Smart Adaptive Systems)、智能代理(Intelligent Agents)、支持向量机(Support Vector Machines),其中智能代理的应用主要是为了使不同版本的用户模型被包含在一个统一的用户模型中。SUM 还采用了机器学习法中的归纳法和演绎法,实现了开放的、分布式的、异构的推荐系统中的个性化服务,用户可以在不同领域中相互交流信息。另外,情感因素^[11]的引入可以更加全面地描述用户的特征。目前,该模型还处于研究阶段,没有实际的应用。

从上面两个实例可以看出,这些方法大多是通过映射技术来融合不同的用户模型,可以是异构的用户模型共存,也可以是把分布式的用户模型实例聚合成一个用户模型。

2.3 小结

构建跨系统的用户模型,关键是实现用户模型的复用。通过分析以上实例,可以从用户特征表示方式、用户模型复用方法、关键技术支撑以及模型的可扩展性和自适应性几个方面进行比较分析,如表 1 所示:

表 1 跨系统用户建模方法特征比较

方法	基于统一-用户模型的方法			基于用户模型间映射和融合的方法	
	OntobUM	UUCM	GUMO	GUC	SUM
用户特征表示方式	本体	本体或词汇表	本体	OWL 和图	属性-值
关键技术支撑	RDF(S)、OWL、本体集成	RDF(S)、OWL、信息抽取	RDF(S)、OWL	匹配融合技术、RDF(S)、OWL	机器学习、映射技术
用户模型复用方法	用户本体、领域本体、日志本体	情境护照	UserML	UAV 知识库	智能代理
可扩展性	一般	很好	一般	一般	很好
自适应性	一般	很好	一般	一般	很好

可以看出,这两类方法都是基于语义网技术的支撑,本体在用户模型的复用中起到了很大的作用。正是基于本体的复用性和共享性,很多系统都试图利用

本体来实现与其他系统之间的交流。第一类方法自上而下,致力于构建一个可以适用于所有系统的用户模型,从而实现跨系统服务;第二类方法自底向上,主要

考虑如何把不同系统间的用户模型映射和融合成一种统一的形式,从而可以在不同系统间相互交换数据。实际上,构建一个可以满足所有系统需要、满足所有情境的通用本体并不容易。尽管这种方法可以实现模型的共享和系统间的交互,然而,现实中很难让系统由原本已有的用户模型向新的共享模型转换。与基于统一用户模型的方法相比,基于用户模型的映射的方法显得比较灵活。然而,由于不同系统中数据的异构性,不同用户模型模式映射过程会比较繁杂。

3 跨系统用户建模的新范式——分布式开放用户建模

在语义网环境下,跨系统的用户建模方法已经由集中控制系统向动态联合的服务系统转变。很多学者对此进行了研究,下面介绍几个典型实例。

3.1 基于 Mypes 的社交网络用户建模

Mypes 提供了一种连接社交网络(如 Facebook、LinkedIn)、社交媒体(如 Flickr、Delicious、Twitter)、Google 等应用中不同用户模型数据的服务,可以对异构的数据进行一致化,并且能够通过 WordNet^[12]来增强用户模型的语义化。Mypes 使用户可以浏览分布在不同系统中的用户模型数据,同时以 RDF 或 vCard 的格式使应用系统对用户模型进行聚合、语义增强。Abel 等^[13]于 2010 年利用 Mypes 聚合了基于表单(Form-based)和基于标签(Tag-based)的用户文档,并且构建了社交网络的用户模型,包括账户映射(Account Mapping)、用户模型数据的聚合(Profile Aggregation)、用户模型数据的对准(Profile Alignment)、语义增强(Semantic Enrichment)几个部件,如图 1 所示:

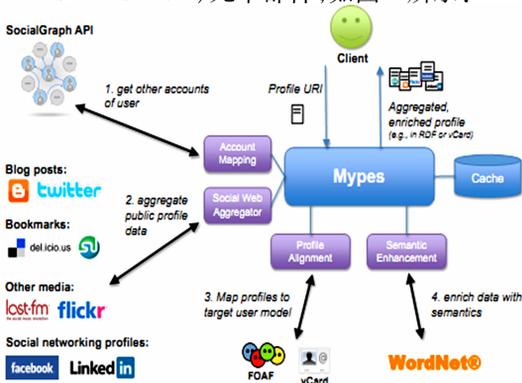


图 1 基于 Mypes 的用户模型数据的聚合和语义增强^[13]

该方法通过对基于表单和基于标签的用户模型数据的聚合和语义增强,涵盖了用户的显式数据和隐式数据。主要采用基于 WordNet 的语义增强,以便对基于标签的用户数据进行很好的归类。该模型的实现还需要多方面的支持: APIs 标准(如 OpenSocial^[14]), 验证与授权协议(如 OpenID^[15]), 开放授权(如 OAuth^[16]), Web 标准(如 RDF, RSS)和具体的微格式(如 hCard^[17]或 RelTag)等。基于 Mypes 的跨系统用户建模方法能够更加全面地捕捉用户的信息,提高用户推荐系统的质量。不同系统中的用户模型数据相互补充,很好地解决了系统的冷启动问题。随着 FOAF^[18]、SIOC^[19]、GUMO^[6]等标准的出现,连接分布式的用户数据变得容易起来。

3.2 分布式互操作用户模型

Carmagnola^[20]于 2009 年提出了用户模型互操作的逻辑架构,利用互操作性很好地解决了分布式用户模型的语义异构问题,如图 2 所示:

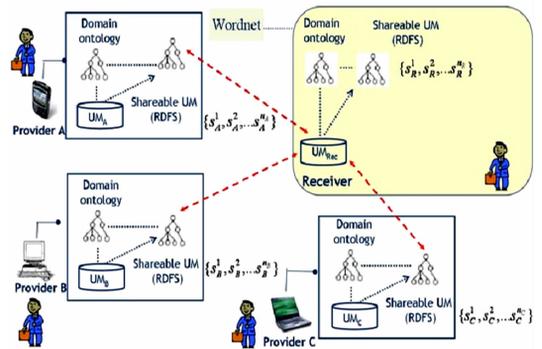


图 2 用户模型互操作的逻辑架构^[20]

每个系统都有一个可共享的用户模型,并且与领域本体相关联。需要用户模型数据的系统(Receiver)从已具有可共享用户模型的系统(Provider)中获得用户数据,再结合特定的领域本体构建属于自己的用户模型。

本方法实现了跨系统的用户建模,通过共享和重用用户数据,采用共享的词汇格式(如 FOAF、vCard 等)来表示用户模型不会产生数据异构的问题。在开放动态的环境下,要求所有的系统采用共用词汇表的方法是不切实际的。该方法采用中间解(Intermediate Solution)和语义映射技术来表达用户模型,具有很好的灵活性。然而,要求每个系统维护一个可共享的用户模型才能进行互操作过程,用户模型数据的任何变化都

需要复制到可共享用户模型中。而且,用户模型的互操作过程需要很多方面的计算能力,导致反应结果的推迟。为了适应动态的变化,系统需要定期地检查与更新用户模型。

3.3 关联数据驱动的用户建模

随着关联开放数据(Linked Open Data, LOD)的产生和应用,用户在 Web 上已不是单一孤立的个体,用户在不同系统之间、不同的用户之间都存在着潜在的关联。目前,在 Web 上已经产生了很多相互关联的用户数据,利用关联数据构建用户模型是一种很好的尝试。2011 年, Ye 等^[21]提出了基于 CUM(Core User Model)的用户建模框架,主要包括概念映射(Concept Mapping)模块、关联(Interlinking)模块、更新(Updating)模块,如图 3 所示:

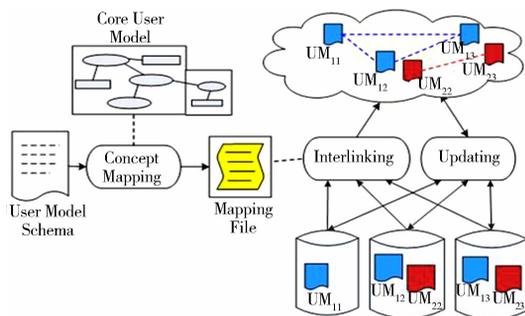


图 3 关联数据驱动的用户建模框架^[21]

其中,用户模型模式(User Model Schema)是对用户模型的抽象描述,定义了用户模型的组成元素。通过概念映射,形成映射文件,包括概念之间的一致性、术语的联合和特定领域的词汇。根据映射文件进行关联,对于少量的数据可以人工完成。大量数据需要自动地完成关联^[22],通常使用基于模式的匹配(Pattern-based Matching)和基于属性的方法(Attribute-based Methods)。该方法采用核心用户模型(Core User Model, CUM)来描述用户共同的属性和特定领域的特征,并以 FOAF 的形式来表示^[21]。CUM 使用 rdf:seeAlso 属性来扩展表达用户的其他特征。

这种方法主要借鉴关联数据的应用,把系统的用户模型按照一定的程序和规则发布成关联数据,不同系统中的同一用户之间就产生了相互关联。这样,关联的用户模型构成了不同应用系统间的跨系统用户模型,可以提供更加全面完整的用户信息。LOD 的应用,不仅允许推断用户之间的关系,还可以在更高层次的

概念上对用户的兴趣进行聚类。目前,该方法还没有具体的应用,但为跨系统用户建模开辟了新的路径。

上述三个实例分别从社交网络、互操作、关联开放数据的角度实现了跨系统的用户建模,都是分布式开放的用户建模方法。主要是把分布在不同系统中的数据收集聚合,或者对不同系统中的用户模型数据进行关联一致化。如何把 Web 上大量分布式的、异构的用户数据进行概念的一致化是分布式开放用户建模的关键,数据开放也是跨系统用户建模面临的首要问题。分布式开放用户建模不仅需要语义网技术的支撑,还需要其他政策、法规等方面的支持。目前,这种方法是语义网环境下用户建模的一种新范式。与前面两类方法相比,本体在分布式开放用户建模中也承担了重要角色,如用户特征的表达、用户关系的表达等。这种方法更加注重数据的开放性和交互性,具有更好的可扩展性和自适应性,试图随时随地构建满足用户需要的模型。可以说,分布式开放用户建模是 Web 发展的必然结果。

4 结 语

随着 Web 的发展,用户面临着越来越严重的信息过载的问题,个性化服务为用户提供了一个很好的出路。跨系统的用户建模可以为用户提供多系统、多应用领域的个性化服务,从而构建更加全面丰富的用户模型,很好地解决了传统用户建模中的数据冗余和冷启动问题。本文比较分析了跨系统用户建模的一些典型实例,在此基础上指出了跨系统用户建模的发展趋势——分布式开放用户建模。社交网络、互操作、关联开放数据的发展和应用,为跨系统的用户建模提供了新的契机。

目前,个性化服务的研究已经从关注用户模型的本身向用户模型的构建过程转变,不同系统之间形成一种动态联合的服务模式。在分布式、开放的 Web 环境中,如何尽可能全面地收集用户信息、如何有效地组织异构数据源、如何实现不同系统间的数据交换和共享、如何快速地适应不同用户的需要、如何高效地捕捉用户的兴趣和爱好是要考虑的重要问题。语义网技术的发展、关联开放数据的发布和应用,为这些问题的探索和研究提供了有效的途径。

参考文献:

- [1] Abel F, Araújo S, Gao Q, et al. Analyzing Cross - system User Modeling on the Social Web[C]. In: *Proceedings of the 11th International Conference on Web Engineering*. 2011;28 - 43.
- [2] Mehta B, Niederée C, Stewart A, et al. Towards Cross - system Personalization[C]. In: *Proceedings of the 3rd International Conference on Universal Access in Human - Computer Interaction*, USA. 2005;119 - 123.
- [3] Viviani M, Bennani N, Egyed - Zsigmond E. A Survey on User Modeling in Multi - application Environments[C]. In: *Proceedings of the 3rd International Conference on Advances in Human - Oriented and Personalized Mechanisms, Technologies and Services (CENTRIC)*. 2010: 111 - 116.
- [4] Razmerita L, Angehrn A, Maedche A. Ontology - based User Modeling for Knowledge Management Systems[C]. In: *Proceedings of the 9th International Conference on User Modeling*. 2003: 213 - 217.
- [5] Niederée C, Stewart A, Mehta B, et al. A Multi - Dimensional, Unified User Model for Cross - system Personalization[C]. In: *Proceedings of the AVI 2004 Workshop on Environments for Personalized Information Access*, Gallipoli, Italy. 2004: 34 - 54.
- [6] Heckmann D, Schwartz T, Brandherm B, et al. GUMO - The General User Model Ontology[C]. In: *Proceedings of the 10th International Conference on User Modeling*. 2005;428 - 432.
- [7] Heckmann D. Introducing "Situational Statements" as an Integrating Data Structure for User Modeling, Context - Awareness and Resource - Adaptive Computing[C]. In: *Proceedings of the 11th GI - Workshop on Adaptivity and User Modeling in Interactive Software Systems (ABIS - 2003)*, Karlsruhe, Germany. 2003;283 - 286.
- [8] Hebel J, Fisher M, Blace R, et al. *Semantic Web Programming* [M]. Wiley, 2009.
- [9] Van Der Sluijs K, Houben G J. Towards a Generic User Model Component[J/OL]. [2012 - 03 - 01]. <http://www.win.tue.nl/persweb/Camera-ready/%2313-Sluijs-full.pdf>.
- [10] González G, López B, De La Rosa J L. A Multi - agent Smart User Model for Cross - domain Recommender Systems[C]. In: *Proceedings of Beyond Personalization: The Next Stage of Recommender Systems Research*. 2005.
- [11] González G, López B, De La Rosa J L. The Emotional Factor: An Innovative Approach to User Modelling for Recommender Systems [J/OL]. [2012 - 03 - 01]. <http://eia.udg.es/~blopez/publicacions/RPeC02.pdf>.
- [12] Miller G A. WordNet: A Lexical Database for English[J]. *Communications of the ACM*, 1995, 38(11):39 - 41.
- [13] Abel F, Herder E, Houben G J, et al. Cross - system User Modeling and Personalization on the Social Web[J]. *User Modeling and User - adapted Interaction: Special Issue on Personalization in Social Web Systems*, 2011, 22(3):1 - 42.
- [14] Mitchell - Wong J, Kowalczyk R, Roshelova A, et al. Opensocial: From Social Networks to Social Ecosystem[C]. In: *Proceedings of Digital EcoSystems and Technologies Conference*. 2007: 361 - 366.
- [15] Recordon D, Reed D. OpenID 2.0: A Platform for User - centric identity Management[C]. In: *Proceedings of the 2nd ACM Workshop on Digital Identity Management*. ACM, 2006: 11 - 16.
- [16] Hammer - Lahav E. The OAuth 1.0 Protocol[EB/OL]. [2012 - 03 - 01]. <http://art.tools.ietf.org/html/rfc5849>.
- [17] Eralai I, Dragulescu B. The Usefulness and Functionality of Microformats in a Particular ELearning System[C]. In: *Proceedings of 2010 International Joint Conference on Computational Cybernetics and Technical Informatics (ICCC - CONTI)*. 2010: 387 - 390.
- [18] Brickley D, Miller L. FOAF Vocabulary Specification 0.91[EB/OL]. [2012 - 03 - 01]. <http://lyle.smu.edu/~coyle/cse7347.prev/handouts/s14.FOAF%20Vocabulary%20Specification.pdf>.
- [19] Breslin J, Bojars U, Passant A, et al. SIOC: Content Exchange and Semantic Interoperability Between Social Networks[EB/OL]. [2012 - 03 - 01]. <http://aran.library.nuigalway.ie/xmlui/handle/10379/629>.
- [20] Carmagnola F. Handling Semantic Heterogeneity in Interoperable Distributed User Models[A]. //Kuflik T, Berkovsky S, Carmagnola F, et al. *Advances in Ubiquitous User Modelling* [M]. Berlin, Heidelberg: Springer - Verlag, 2009;20 - 36.
- [21] Ye N, Zhao Y L, Zhao X, et al. A Linked Data - driven Solution to User Modeling in a Multi - application Environment[C]. In: *Proceedings of the 2nd International Conference on Software Engineering and Service Science (ICSESS)*. IEEE, 2011: 872 - 876.
- [22] Bizer C, Cyganiak R, Heath T. How to Publish Linked Data on the Web[J]. *Computer and Information Science*, 2007, 20:2008.
- (作者 E - mail: niuyazhen10@163.com)