

主题图的概况及应用研究* (上篇)

□ 李颖 练霞 朱礼军 / 中国科学技术信息研究所 北京 100038

刘静波 / 日本筑波大学体育研究科

摘要: 在语义Web时代,主题图(Topic Maps,文章简称TM)被认为是实用性较高的知识组织技术。本研究以TM技术的应用为最终目标,分上、下两篇整体地论述TM的相关研发工作。本篇中给出TM的概况,包括TM的概念、标准体系、应用领域、开发工具和相关网站等,为进行具体领域的应用开发奠定了基础;在另一篇中,将结合金融培训机构知识体系,介绍如何应用TM技术来组织行业知识,包括基于TM技术的金融培训机构知识系统的架构设计和实现。最后,给出本研究对未来的展望及结论。

关键词: 主题图, TM, ISO/IEC 13250, 知识组织, RDF, 语义Web, 金融培训

DOI: 10.3772/j.issn.1673-2286.2010.05.005

序言: TM, 面向信息、知识间的无缝链接

TM是信息资源之上,即Meta层次(Meta Layer)上的语义化信息组织、检索、导航规范。追溯TM的起源,1991年Davenport研究小组启动了TM工作,1993年在其工作文档中,率先使用了TM一词;其后,经过HyTime^①成员的一系列活动,自1995年起,TM成为ISO持续研讨的标准化项目;2000年,TM上升为ISO13250这一国际标准。具体地说,TM是由ISO/IEC JTC1 SC34制定的标准,目前该标准还在继续完善。在国际性非盈利联盟OASIS中,也在不断地研讨制定一系列TM相关标准(详见本文

第3部分)。就是说,TM是全球化信息组织的标准技术^[1]。

从信息组织方法分类来看, TM源于图书索引,它提供图书目录之外可检索图书信息资源的手段。而数字信息资源的TM,就如同数字信息空间的“导航仪”。TM在数字信息空间中与信息进行分类、体系化,从而利于信息资源的发现。它针对信息空间问题域中的主题、主题间的关系,以及主题与信息资源间的关系,通过Topic、Association及Occurrence三要素来进行模型化,实现计算机的处理。

TM与传统的信息组织方法的不同点是,它集各家之长,可以描述主题概念间从简单到复杂的任意

关系,进行语义处理。就是说, TM在语义Web时代,实现了导航用户访问必要的资源内容,是面向信息、知识间无缝链接的技术。

1 本研究的背景与动机

伴随着TM在语义Web时代实用价值的凸显,国外在众多的行业领域已经开发构建了基于TM的知识系统。在这一背景之下,国内TM理论与应用研究开始受到了瞩目,原型系统的开发也有所发表,但这些还远远不到实用化程度。如图1国内TM综述性研究知识图所示,目前有关TM的研究,从概念的角度,已经涉及到了从基础理论、技术工具、实际应用,到国内主要研

* 基金项目: 国家“十一五”科技支撑计划课题“知识组织系统的集成及服务体系研究与实现”(2006BAH03B03)和“科技文献信息服务系统关键技术研究及应用示范”(2006BAH03B06)资助项目。中国科学技术信息研究所重点项目“汉语科技词系统建设与应用工程——新能源汽车领域完善及领域扩展”(2009KP01-3-2)资金项目。全XML流程多样性出版服务研究。
^①即Hypermedia/Time-based Structuring Language, 超媒体/基于时间的结构语言,是超媒体信息结构化表示的国际标准(ISO/IEC 10744),是通用标记语言(SGML, standardized general markup language)的一个应用。

究、现状（存在问题）、研究方向建议等方面^[2]。这说明国内的研究者已经充分地意识到了TM的实用价值，并在开始寻求融合TM技术的语义化信息与知识组织相关解决方案。

笔者自2002年开始接触TM技术，有幸在东京倾听了TM开拓人之一Lars Marius Garshol在研讨会上的发言，对TM技术在信息导航中的作用颇有感悟。2003年，尝试开发了基于TM技术的科技信息资源检索系统的导航功能，增强了系统的检索功能^[3]。近年，为了发挥TM在语义处理上的特长，我们在继续探索将进行语义处理的TM技术，融入到现有的知识组织体系中，研究语义Web环境下，知识组织的最佳实践方案。近期尝试了

在金融培训机构知识体系中，运用TM技术，构建具有知识分类和知识导航的系统，并试图向其他领域扩展^[4]。

为了与国内同行分享我们的研究经验，较全面地反映出整个研究过程，本研究将分为上下两篇文章，有的放矢地介绍TM理论体系和实践开发的流程。本文作为上篇，首先采用与其他信息组织方法对比的方式来理解TM的概念；通过与语义处理RDF技术的对比，把握TM的标准体系的发展；在此基础上，综合性地给出TM的适用领域、开发工具和相关网站。下篇将从系统开发的角度，介绍基于TM的金融领域远程教育知识体系的架构设计与开发流程。

作为信息与知识组织的方法之一，TM与人们所熟知的手段有着密切的关联。为了更好地理解TM，不妨可以将TM与其他相关方法“相提并论”，进行比较，以示区别。基于目前的研究趋势，可以简单地将信息组织方式分为基于元数据和主题两种类型，TM可以看作基于主题信息组织中的一种，各种方法的要点如下：

■ 基于元数据：关于数据的数据，或称关于文档及Web页等信息资源（对象）的数据，比如Dublin Core。

■ 基于主题的分类法：

①. Taxonomies：主题（概念）的阶层分类。主题间的关系只有广义-狭义。

②. Thesauri：有关的国际标准为ISO2788、ISO5964。主题（概念）的阶层分类，除了广义（BT）-狭义（NT）的关系，还有：词的语义说明Scope Note；关联词RT；反义词AT（antonym）；同义词SY；优先词USE；非优先词UF；最上位词TT。

③. Faceted classification：依据复数个面进行分类，比如可以基于“personality、matter、energy、space、time”5个面分类。

④. Topic Maps：主题（Topics）与关联（Associations）的集合；主题间的关联关系可自由定义；可链接到有关主题的信息资源；具有唯一识别主题的机制，比如URI；可设定有效的范围（Scope）；具有合并功能；容易具体化。

⑤. Ontology：本体的本质是从对象世界中分离出基本概念，形成概念的集合。概念间除了上下位关系（is-a关系）即阶层关系以

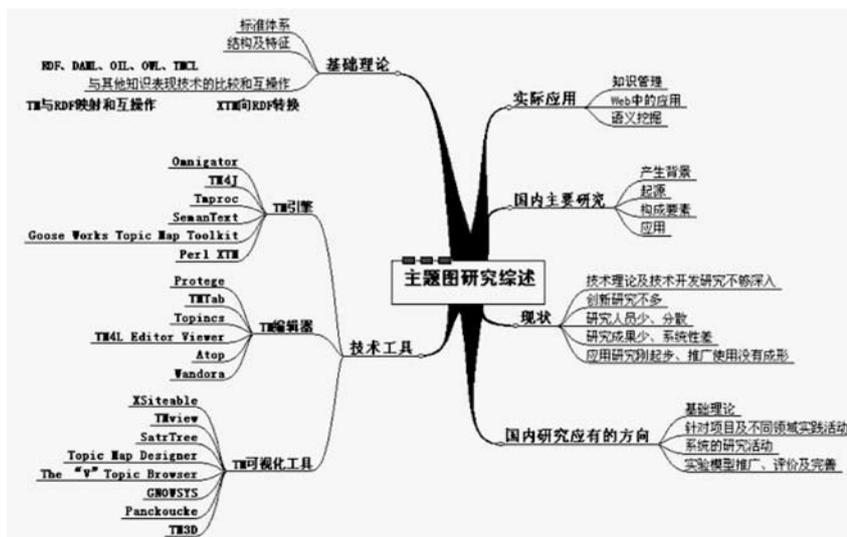


图1 国内TM综述性研究知识图

2 TM的概念

TM既吸收了传统图书馆信息学组织手段的各家之长，又吸收了人工智能的思想，必然与这些方法有着密切的关联与不同，所以有

必要从多个视角来解析和明确TM的概念。

2.1 TM在信息组织方法中的定位

外，还具有其他必要的关系。即，本体抽出概念和关系的定义，对语义限制公理化。

从以上的对比看出，TM的优势是，在处理语义方面，比本体更简单易行，与人的思维方式更接近，容易具体化。从以下TM的定义中，可以具体体会TM的这一优势。

2.2 TM的定义

TM的思想非常质朴，可以追溯到1923年Ogden & Richards^[5]的语义三角理论。如图2所示，语义三角反映了现实世界的对象（Object）、概念（主题，Subject）与记号（主题，Topic）之间的关系^②。

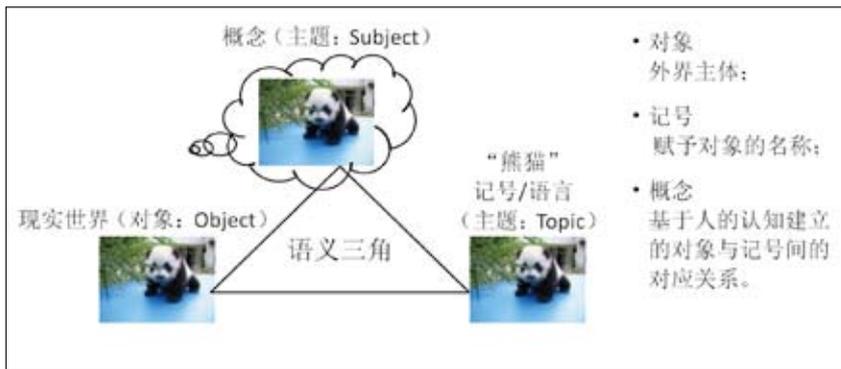


图2 对象、概念（主题：Subject）、记号（主题：Topic）之间的关系

TM又是如何进化的？参见图3主题、关联、出现之间的关系。TM是将信息资源的主题、主题间的关系，以及主题与信息间的关系，利用Topic（主题）、

Association（关联）以及Occurrence（出现）3种构成要素来模型化，实现计算机的处理。主题、关联及出现是TM构成的三大要素，取它们英文单词的首字母构成的TAO

of Topic Maps，TAO正好是中文的“道”的拼音，TM技术核心就是主题图的“道”。TAO的各要素寓意是：

■ 主题（Topic）：人类认识的具体或抽象的主题与概念；

■ 关联（Association）：关联用来表现主题之间的关系；

■ 出现（Occurrence）：指向与主题相关的信息资源。

图3还显示出了TM具有3种指针：

■ topicRef：指向topic的指针；

■ subjectIndicatorRef：指向表示主题所在信息资源的指针；

■ resourceRef：指向构成主题信息资源的指针。

TM的关联表现概念间的关系。比如，在XTM（基于XML语法的TM）标准中，用class-instance关系定义topic间的类和实例的关系，用superclass-subclass定义topic间的上位类和下位类的关系。此外，关联可以自由定义下述概念间的任意关系：等价关系（同义、类似、对比），阶层关系（上位-下位、全体-部分），联想关系（并列、用途和环境、因果），及其他任意关系。

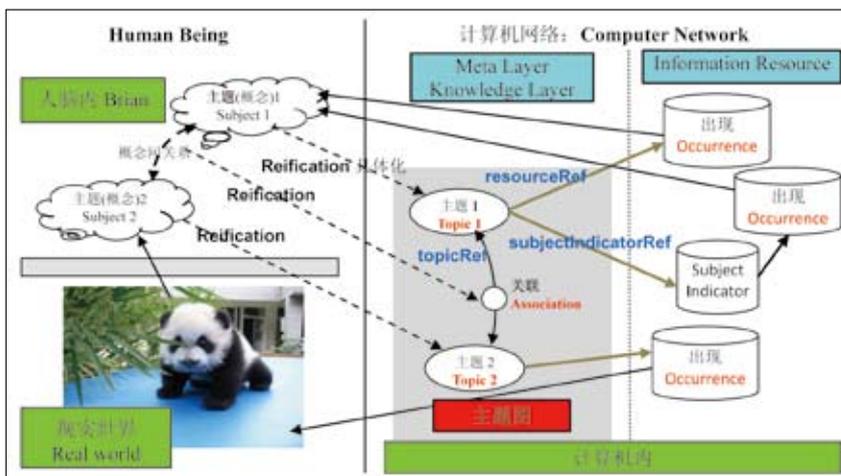


图3 TM中主题（Topic）、关联（Association）及出现（Occurrence）

2.3 TM的特点

■ 概念、语义级别的信息操作；

■ 信息实体与元层（Meta Layer）的分离：

◆ 不需要变更信息实体就可进行信息的组织化、体系化；

◆ 概念体系元层的模型化较

^②注释：概念世界的主题Subject与记号世界里（或计算机内）的主题Topic英语是不同的词汇，所以，有个别学者将TM中的Topic译成主题，而将Subject译成标题，以示区分。本文有时提及的主题并不需要区分它是概念领域还是计算机内的主题，所以我们将Subject和Topic都翻译成主题。而特指概念世界的主题时，对应的英语是Subject，论及计算机内处理的主题，对应的英语是Topic，请根据上下文的场景，予以区分，以免混淆。

为容易;

◆有利于数字内容的分类、

管理。

■ 可以对主题进行严密的识别;

■ 直观且易熟悉;

■ 基于标准:

◆信息资源的永久性确保;

◆多种工具可选择和利用;

◆平台、提供商的独立性。

■ 对信息资源多样性观点的设定和基于不同视点导航;

■ 可依据利用者个人视点

(Personalization) 进行信息的组织。

3 TM标准化活动及主要标准

3.1 ISO/IEC JTC1 SC34

TM标准制定工作由ISO/IEC JTC1 SC34 (文献描述和处理语言)的WG3 (信息关联化工作组)进行。国际会议与美国及欧洲举办的XML会议等并设,大体每年举办4次会议。ISO国际会议与W3C、IETF及OASIS不同,参加成员仅限于国家实体。WG3的会议召集人(convener)为挪威的Steve Pepper,主要的参与国家有美国、法国、英国、挪威、日本,以及德国等。

早在2001年12月17日,ISO/IEC JTC 1/SC34 就公布了Roadmap of the further work on topic maps^[6],如图4所示,ISO有关TM的标准化工作路线图如下:

(1) 参考模型RM (Reference Model): 基于PMTM4 (Processing Model for XTM 1.0) 的文档,定义

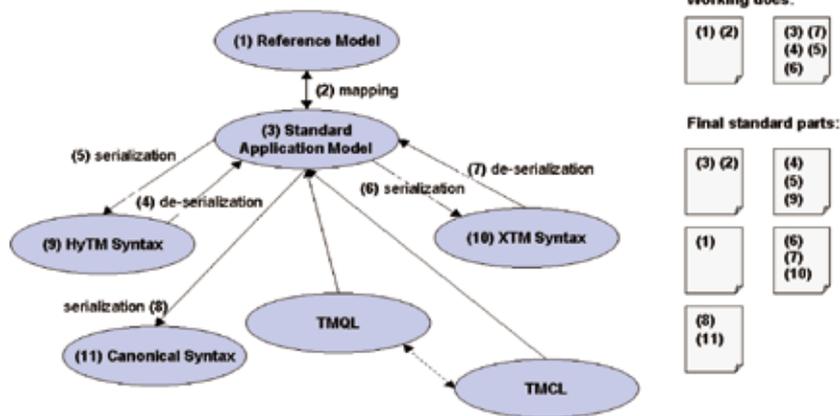


图4 主题图标准体系发展路线图

主题图的参考模型,定义TM与其他知识表现的关系。

(2) 用参考模型RM定义标准应用模型SAM: 作为参考模型的应用,定义TM的标准应用模型SAM (Standard Application Model)。

(3) 主题图标准应用模型: 基于既有信息集 (infoset) 方案的主题图模型。

(4) XTM的系列规范: 基于XTM 1.0文档如何构建SAM实例的细节性形式化规范。

(5) XTM系列推荐: 是关于如何序列化SAM实例为XTM 1.0语法的形式化描述。

(6) HyTM系列推荐: 是关于如何序列化SAM实例为HyTM的形式化描述。

(7) HyTM系列化规范: 是关于如何从HyTM文档构建SAM实例的详细的形式化规范。

(8) 权威的规范: 是关于如何序列化SAM实例为规范化主题图文档的详细的形式化规范。

(9) HyTM语法规则: 对既有的HyTM语法在措词上进行改写,以适应新的标准框架。

(10) XTM 1.0的语法规则: 对

既有的XTML语法在措词上进行改写,以适应新的标准框架。

(11) 权威的语法规则: 权威的SAM实例的语法规则。过程描述如(8)。从SAM实例中产生这一语法的规范。

3.2 OASIS

OASIS的TM相关标准化活动是与Published Subject (主题发布)有关的3个委员会,主要进行Published Subject标准化及普及工作。Published Subject定义了可共同利用的主题 (Topic), 在Web上永久公开。不仅TM之间,与其他知识表现语言之间,均可容易地实现知识的共享和交换。有关的国际会议与SC34 WG3同样,与美国及欧洲举办的XML会议等并设,大体每年召开4次。主要成员也是SC34 WG3的成员。

3.3 TM相关的主要标准

由于TM发展在地域上的不平衡性,参与TM国际标准制定的国家非常有限,并且至今为止,我国

应用TM技术的需求也不够强烈，所以对TM国际标准进行系统化跟踪与研究的工作做得还很不到位。为了介绍TM标准的发展历史，本文引用TM标准制定主要参与人之一，挪威出生的Lars Marius Garshol先生绘出的TM发展年表（与RDF对照）来纵览TM的变迁。如图5所

示，伴随着RDF向本体语言OWL方向进一步发展，融合了TM语义处理优势的知识系统的实战研发，给语义Web带来了新的活力。对比较为人熟知的RDF标准体系来介绍TM显得很有必要，可以体现出信息技术融合的协同效果（请参见本文展望部分的有关内容）。

言。

■ TMCL：主题图用Schema语言。

■ LTM：Linear Topic Map Notation，线性主题图记法。利用普通的文本编辑器就可以简单地写出主题图格式文档，便于个人书写简单的主题图文档、通过电子信函交换部分主题图。此格式可以转化为XTM。

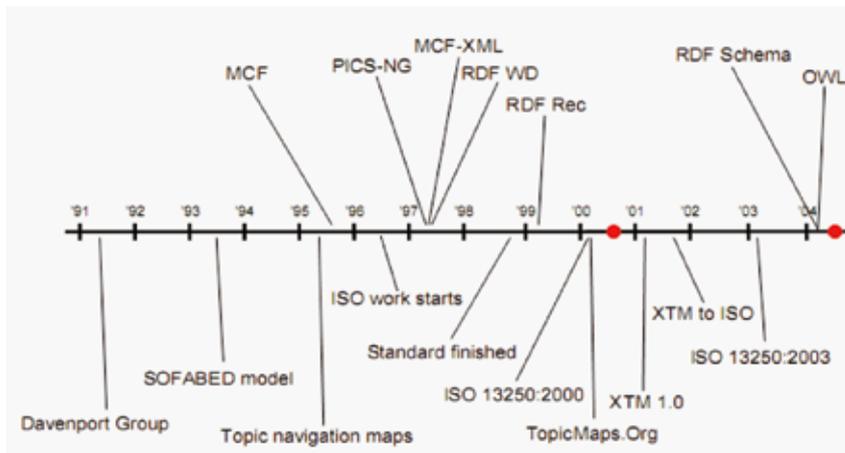


图5 TM发展年表（与RDF对照）

[出自：TM/RDF Interoperability in Practice, Lars Marius Garshol]

图6是TM相关的主要标准，通过与RDF标准体系的对照，可以看出：TM与RDF在标准体系上，有着很多的相同点。正因为如此，TM与RDF的融合是大势所趋。TM主要标准如下：

■ HyTM：HyTime的Topic Maps，基于HyTime的语法。

■ XTM：XML Topic Maps。规定了创建主题图的XML标签集和相应的语法。

■ TMQL：主题图用查询语

4 Topic Maps应用领域与案例

4.1 国外主要应用研究与案例

■ fuzzy.com (<http://fuzzy.com/>)：是Web科学与Web开发社团。它与普通的社交标签和网站不同，其标签及其标签之间的关系是基于TM的语义处理。

■ 主题图应用案例：日本基于TM的最权威的知识协同门户。
<http://www.knowledge-synergy.com/openspace/tmcase/presentation-list.html>

■ 2007年召开的3个主要TM国际会议发表的研究文章

- Topic Maps 2007, Oslo, Norway, 2007.3.20-21 <http://www.topicmaps.com/tmc/conference.jsp?conf=TM2007>

- TMRA 2007, Leipzig, Germany, 2007.10.10-12

<http://www.informatik.uni-leipzig.de/~tmra/2007/>

- AToMS 2007, Kyoto, Japan, 2007.12.12

<http://www.knowledge-synergy.com/news/atoms2007-en.html>

■ 其他领域的案例

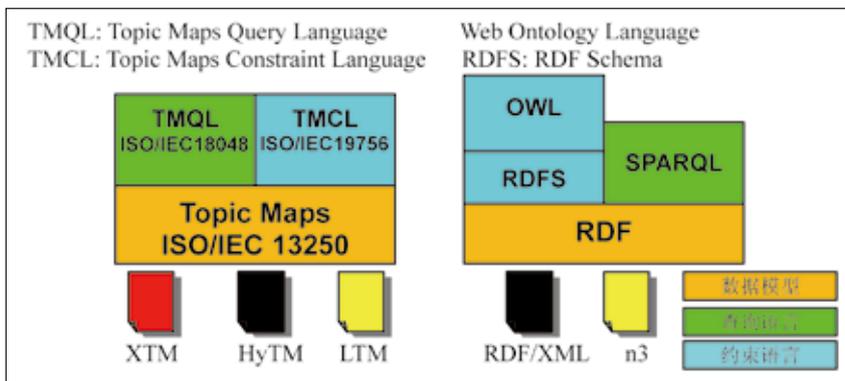


图6 TM主要标准（与RDF对照）

- 出版 (电子出版、印刷物)
 - ◆ 美国 美国财政部国家税务局 (IRS) 有关税务的出版物
 - ◆ 荷兰 税务局 为纳税人提供信息
 - ◆ 法国 百科全书 (quid)
 - ◆ 挪威 百科全书
 - ◆ XML Conference Proceedings
 - Web应用开发
 - ◆ 挪威 ITU教育项目用Web网站
 - ◆ 南非 CSIR iWorks Ideabank研究所的知识/构想的共享
 - ◆ Patrimoine金融关系文档用的内部网
 - 应用开发
 - ◆ Bravo知识管理工具
 - 应用统合
 - ◆ StarTeam软件开发/维护的协同作业环境
 - 内容管理
 - ◆ 美国 能源部 机密文档分类指南/管理
 - e-Learning、telelearning、CAL (Computer Aided Learning)
 - ◆ 德国 教材的导航
 - ◆ 挪威 BrainBank 学习支持系统
 - Web门户
 - ◆ 挪威 面向年轻人的科技信息门户
 - ◆ 挪威 消费者信息的门户
 - ◆ 挪威 农业部门户
 - ◆ 挪威 保守党门户
 - ◆ 挪威 司法部门户
 - ◆ 挪威 税务署门户
 - ◆ 挪威 首相官邸门户
 - ◆ 挪威 统计局门户
 - ◆ 挪威 原子炉项目门户
 - ◆ 挪威 文化部门户

- 电子调拨、贸易交易
 - ◆ 韩国 ETRI Collaborative Product Commerce (CPC) 项目
 - ◆ 电子目录的产品分类体系的标准UNSPSC (Universal Standard Products and Services Classification Code) 主题图
 - 商务流程模型
 - ◆ Shell石油 (Topic Maps和NewsML的并用)
 - ◆ 欧洲议会
 - 其他
 - ◆ 政府机构、制药公司、标致汽车、……

4.2 国内应用研究 – 金融知识体系组织^{[2][4]} (参见本研究的下篇)

5 TM的开源工具和有关网站

5.1 TM开源工具和有关网站

5.1.1 开源工具:

- Omnigator www.ontopia.net
- TM4J tm4j.org
- Seman Text www.semantext.com
- Nexist nexist.sourcefoge.net
- TMTab www.techquila.tmtab.html

5.1.2 有关网站:

- 主题图Web 应用 (Topic Maps Web application)
 - Everyday Physics on Web (EPW): http://tm.u-gakugei.ac.jp:8080/epw/

- Topic Map for Topic Maps case examples (主题图应用实例主题图): http://www.garshol.priv.no/tmcase1/
- Lars Marius's photos: http://www.garshol.priv.no/tmphoto/
- Topic Maps tools: http://www.garshol.priv.no/tmtools/
- fuzzy.com: http://www.fuzzy.com/tag/?id=2238
- 主题标识符 (Subject Identifier)
 - ONTOPIEDIA's PSI Server: http://psi.ontopedia.net/
 - subj3ct: https://subj3ct.com/
 - DBpedia: http://dbpedia.org/About
 - 标准、项目、讨论会等 (Standard, Project, Conference, etc.)
 - Ontopia project: http://code.google.com/p/ontopia/
 - Ontopia's Blog: http://ontopia.wordpress.com/
 - TopicMaps.jp: http://topicmaps.jp/blog/
 - topicmaps.community: http://www.topicmaps.com/tmc/conference.jsp?conf=TM2010
 - TMRA 2009: http://www.tmra.de/2009/
 - TMRA: http://www.tmra.de/2008/
 - topicmap.com: http://topicmap.com/
 - ISO/IEC SC34 WG3: http://www.isotopicmaps.org/
 - TopicMaps.Org: http://www.topicmaps.org/
 - Topicmaps.net: http://www.topicmaps.net/
 - xmluk.org: http://www.xmluk.org/public/duxford_2003.html

■ 知识的Concierge: <http://www.kn-concierge.com/concierge/>

■ TM4J project: <http://tm4j.sourceforge.net/>

■ TMAPI: <http://www.tmapi.org/>

■ techquila's World Topic Map: <http://www.techquila.com/topicmaps/tmworld/>

■ TM4L: <http://compsci.wssu.edu/iis/NSDL/>

■ xtm4xmldb: <http://sourceforge.net/projects/xtm4xmldb/>

■ TMwiki: <http://www.topic-maps.org/doku.php>

■ bibMap: http://www.topic-maps.org/cgi-bin/tmv_graph.pl?id=b1&path=bibmap.xtm

■ NZETC: New Zealand Electronic Text Centre: <http://www.nzetc.org/>

■ Seman Text project: <http://www.semantext.com/>

■ Nexist project: <http://nexist.sourceforge.net/>

■ Bond University Topic Maps: <http://topicmaps.it.bond.edu.au/>

■ Bond University Topic Maps Weblog: <http://kill.devc.at/taxonomy/term/22>

■ Easy Topic Maps: <http://easytopicmaps.com/>

■ Robin Cover's Topic Maps pages: <http://xml.coverpages.org/topicMaps.html>

大学、研究机构、公司等 (University, Research institute,

Company, etc.)

■ Ontopia: <http://www.ontopia.net/>

■ Ontopedia: <http://www.ontopedia.net/>

■ Topic Maps Lab: <http://www.topicmapslab.de/>

■ Disclosure Innovation Inc. (日本): <http://tml.synergy.co.jp/>

■ eAid (日本): <http://www.eaid.jp/>

■ NetworkedPlanet: <http://www.networkedplanet.com/>

■ techquila: <http://www.techquila.com/>

■ Cerpus: <http://cerpus.com/>

■ Morpheus: <http://www.mssm.nl/en>

■ Mondeca: <http://www.mondeca.com/>

■ empolis: <http://www.empolis.com/>

■ moresophy: <http://www.moresophy.com/>

■ Jubik: http://www.jubik.com/en_index.html

■ Ligent: <http://www.ligent.net/>

重要人物 (Personal)

■ Lars Marius Garshol's pages: <http://www.garshol.priv.no/>

■ Lars Marius Garshol's Topic map links: <http://www.garshol.priv.no/download/tmlinks.html>

■ Larsblog: <http://www.garshol.priv.no/blog/>

6 展望与结论: TM与其他技术的融合, 实现信

息与知识之间的无缝链接

TM是用于表示知识、进行知识交换的一个标准, 重点在信息的可发现性。TM广义上说是语义网的一种形式, 目前在与相关的语义Web标准进行协调或互操作, 如W3C的系列标准RDF/OWL/SPARQL等, ISO/IEC对TM的标准也在不断地修改, 以便各种技术的融合。

如图7所示, TM与其他技术融合, 实现信息与知识间的无缝链接, 是目前对TM的期待。在图示的TM与RDF标准相互完善和互操作之下, 可以实现多个站点上信息和知识的无缝链接和统合。实现这一目标的技术要素可以包含: 基于TM和RDF的语义结构化数据、任意概念(主题)的全球定位机制Published Subjects、体系化分类概念(知识)的本体、TM片段(fragment)交换(Topic Maps Remote Access Protocol, TMRAP)、查询机制(Query Language)等等。

TM是非常有益的技术, 它为语义Web带来了高度精准且直观的导航检索特性。当然, 对不熟悉TM规范中定义的名词和细节的利用者来说, 这些模型理解起来需要时间。为此, 我们建议通过与人们所熟知的语义技术, 比如与RDF对比来研究TM, 同时, 将TM和RDF进行融合与互通, 一起实现语义Web环境下的信息和知识的无缝链接。

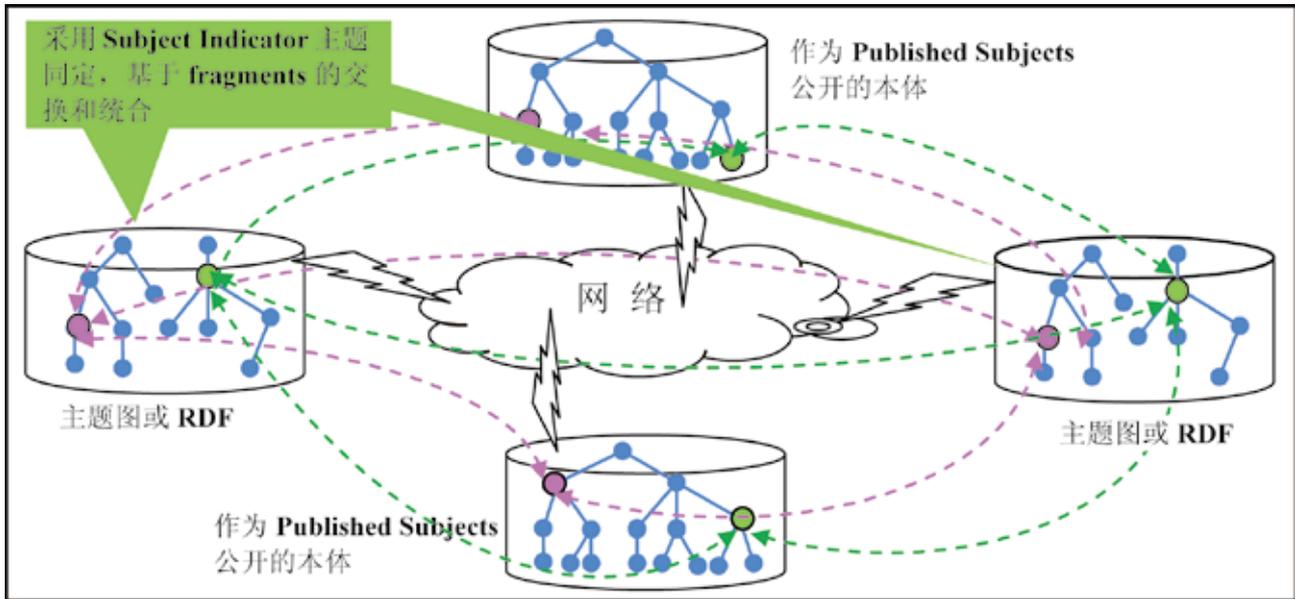


图7 Seamless Knowledge的实现

参考文献

- [1] XTM[EB/OL]. [2009-10-10]. <http://www.topicmaps.org/xtm/1.0/index.html>.
- [2] 刘丹,包平. 国外主题图研究综述[J]. 现代图书情报技术,2007(12):39-44.
- [3] 李颖,石塚英弘.XMLに基づくコンピューター処理可能な構造化デジタルオブジェクトの概念によるWebリソースの検索システムの研究と開発[J]. 情报知识学会社,2003(1).
- [4] 练霞,李颖,等. 构建基于主题图的金融培训机构知识系统(本文part 2/2).
- [5] OGDEN C K, RICHARDS I A. The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism. London: Routledge & Kegan Paul, 1923.
- [6] Topic maps, roadmap for further Work[EB/OL]. [2010-01-17]. <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0278.htm>.

作者简介

李颖, 国立日本筑波大学信息学博士。主要研究方向是: Web的信息知识系统、基于XML的跨媒体数字出版。最近研究课题: 基本XML的数字出版、基于DOI的文献连接系统、跨语言检索、专利分析等。通讯地址: 北京市海淀区复兴路15号 中国科学技术信息研究所 100038。E-mail: liying@istic.ac.cn
刘静波, 国立日本筑波大学体育研究科硕士。主要研究方向: 体育教育课程设置的战术模式与评价, 体育教育课程相关的网络资源系统。最近主要研究方向是: 以体育教育资源为主的信息资源系统。E-mail: seiharyu@hotmail.com

A Survey of Topic Maps and Its application Study (Part 1/2)

Li Ying, Lian Xia, Zhu Lijun / Institute of Scientific and Technical Information of China, Beijing, 100038
Liu Jingbo / University of Tsukuba, Japan

Abstract: In the Semantic Web era, Topic Maps (Hereafter abbreviated as TM) is considered a practical knowledge organization technique. In this study, aiming at the realization of TM technology applications, the two articles give a comprehensive exposition of TM related R&D works. The first article gives an overview of TM, including the concept of TM, the standard systems, applications, development tools and related sites. It lays the foundation for application development in specific areas. The second article, with knowledge system of financial industry training institutions, describes how to use TM technology to organize industry knowledge, including the architecture design of financial industry training knowledge system and its implementation based on TM. Finally, it gives the future prospects and conclusions of the study.

Keywords: Topic maps, TM, ISO/IEC 13250, Knowledge organization, RDF, Semantic Web, Financial industry training

(收稿日期: 2010-04-19)