

●吕元智 王心裁 谭必勇

基于主题地图的电子政务 信息资源组织研究*

摘要 主题地图是一种新型的描述知识结构、知识与信息资源关联的 ISO 标准。从知识层面组织电子政务信息资源,主要就是构建电子政务信息资源的主题地图。步骤是:构建实现模型、准备数据资源、生成主题地图并进行校验。图 2。参考文献 17。

关键词 主题地图 电子政务 信息资源 信息组织

分类号 G253

ABSTRACT Topic map is a new ISO standard for the description of knowledge structures and the relationship between knowledge and information resources. To organize e-government information resources at the knowledge level is to construct topic maps of e-government information resources. There are three steps: constructing realizing models, preparing data resources, generating and testing topic maps. 2 figs. 17 refs.

KEY WORDS Topic maps. E-government. Information resource. Information organization.

CLASS NUMBER G253

电子政务网站上的信息数量、类型虽然很多,但其组织和揭示却非常无序,无法让社会公众准确获取所需信息。由于缺乏有效的信息组织和揭示工具,电子政务网络平台往往成了一个“信息很多,知识很少”的知识盲区^[1]。如何精确组织、揭示、检索海量政府信息,成了当前电子政务工程中一个亟待解决的问题。而新型的信息和知识组织与揭示方法——主题地图,是当前电子政务信息资源组织的一种有效方法。本文试图从主题地图的角度对电子政务信息资源组织进行探讨,以供相关建设和管理部门参考。

1 主题地图

主题地图是国际标准化组织一个用来描述知识结构并把知识同信息资源相关联起来的国际标准(ISO/IEC13250)^[2]。它定义了一个多维的主题空间。在这个空间里,各个位置点表示不同的主题,依据从一个主题到另一个主题所需要转换主题的次数,就可以推算出任意两个主题间的距离,而它们之间的转换关系定义了一个主题到另外一个主题的路径,而路径则用转换的主题来表示^[2]。主题地图实际上是一个知识网络,网络的各个节点代表不同主题的知识,节点之间的连线则代表主题间的联系。根据节点间的连线的数量就可以计算

出任意两个节点间的距离,而这些连线则代表了从一个节点到达另一个节点的路径^[3]。

主题地图的这种多维空间在结构上大致可以分为主题、关联、事件 3 个部分,简称 TAO(Topic, Association, Occurrence)^[4-5]。其中主题一般是根据具体的应用领域和用户需求,对各种不同形态的资源进行主题内容分析后确定的,它可以是人、实体、概念或者其他东西,也即是能引起使用者讨论的对象都可以是一个主题;事件是指在一个或多个层面上与某一主题相关的信息资源。实际使用过程中,事件通常是指储存在于全球信息网络里的任何形式的资源,也即可以通过统一资源定位符(URL)存取到的资源。关联,即主题之间的连线,它表示主题之间的互相关系,类似于传统主题索引中的“参见”关系等^[6]。另外,从信息与知识区分的角度,主题地图的这种结构还可以分为上、下两层^[7]——信息资源集合层(下层)和知识地图(上层)。信息资源集合层也即内容层,它由构成事件的基本信息资源组成,知识地图则由主题及主题之间的关联构成,并且这两层是可以相对独立的。

2 主题地图在电子政务信息资源组织中的意义

主题地图作为“信息宇宙中的全球定位系统”^[8],

* 本文得到了教育部人文社会科学重点研究基地重大项目“电子政务中政府信息资源开发及其技术实现研究”(项目批准号 05JJD870158)资助及上海市教委科研项目(项目批准号 CW0537)资助。

用来对海量的电子政务信息资源进行组织具有十分重要的现实意义。

它可以帮助用户提高信息搜寻能力,方便用户准确地查找到自己需要的政务信息。主题地图在政府信息资源组织方面提供了全新的理念,它从信息内容(知识元)出发来组织信息,并通过关联关系把不同空间位置的信息有机地联系起来,形成非线性的政府信息资源组织结构,提供符合人们思维习惯信息获取模式。主题地图具有清晰的信息导航,可以帮助用户克服需求模糊的局限,从任意一个节点(主题知识)同其他节点的连路找到其他信息,提升信息搜寻能力,准确地获取自己所需要的信息。

它可以有效地实现知识组织,实施政府知识管理,并创造出新的知识,使政府管理更科学。一方面,主题地图是以主题基本概念从知识的层面来组织和管理政府信息资源的,在一定程度上,通过提炼主题等措施将知识资源编码化,将政府信息资源转变成可方便利用的显性知识资源,方便社会公众和政府相关工作者把这些显性知识转为自身的隐性知识,实现知识的转化,实现政府信息资源管理的质的飞跃,即把政府信息资源管理提升到知识管理层面上来。另一方面,在政府信息资源的主题组织过程中,通过主题关联形成多维空间的、组合化的信息、知识结构,其实也是一个新的信息、新的知识产生的过程,是知识创新的过程。主题地图是政府信息资源组织和实施政府知识管理的重要工具,它在实现政府知识资源管理、促进知识创新和改善政府管理等方面有十分重要的作用。

它可以有效地克服政府信息资源孤岛现象,实现信息资源共享,让政府信息资源发挥更大效用。主题地图通过关联,把不同空间的可以搜索到的主题信息组织起来,基于主题地图的政府信息资源组织而形成的政府信息资源结构,它拓宽政府信息资源利用的范围和途径,方便信息用户从一切可能的主题入手搜寻到自己需要的信息。

它拓展了政府信息组织的理论内容,丰富信息组织理论体系。信息的利用得益于信息的组织。如何组织政府信息资源为社会公众提供方便利用是当前政府信息资源组织的重要研究内容之一。由于理念和技术的限制,传统的政府信息组织大部分是按照线性组织模式进行的,不符合人类的思维习惯。主题地图是一种有效的非线性信息组织模式,是按知识组织的要求,通过主题关联关系把相关的信息资源组织起来,是政府信息资源组织的一种有益探索。在很大程度上,主题地图拓展了政府信息资源组织的方式方法,丰富了现代信息组织理论,是值得深入研究的新领域。

3 基于主题地图的电子政务信息资源组织

3.1 构建电子政务信息资源主题地图实现模型

本文构建的电子政务信息资源主题地图实现模型采用三层结构模式,如图1^[9]所示。其中底层为数据层,它由各种形态(如XML文档、网页文档、PDF文档、DOC文档、数据库文件等)的政府信息资源构成。该层主要功能是为电子政务信息资源主题地图生成准备、整理好可供利用的数据。中间层为主题地图生成层,它由主题地图构建环节初步生成主题地图,并交由主题地图的测试、验证环节进行校验,最终形成有效的电子政务信息资源主题地图。上层为用户层,它由主题地图浏览、查询等部分与政府门户网站构成,其主要功能是实现电子政务信息资源主题地图的应用。

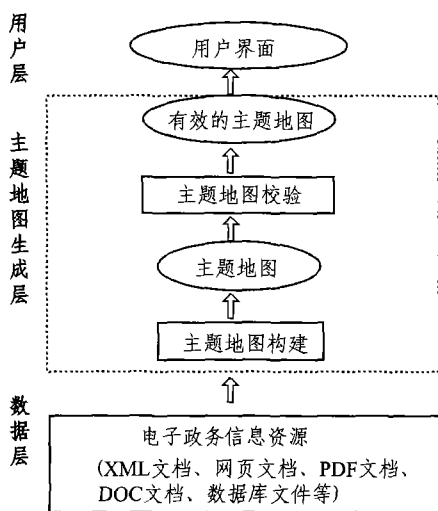


图1 电子政务信息资源主题地图实现模型

3.2 电子政务信息资源的准备工作

信息资源准备工作是电子政务信息资源主题地图生成的基础,主要是尽可能收集和整理符合社会利用需求的政府信息资源,为电子政务信息资源主题地图形成奠定信息基础。它有两个环节。

(1)确定社会对电子政务信息资源的需求。由于人力、物力和资金等条件的限制,政府信息服务部门不可能短时间内对所有的政府信息资源进行有效组织。因此有必要对政务信息资源需求情况进行分析,掌握用户对电子政务信息资源的基本需求。首先,对用户的政务信息资源利用内容情况进行统计分析。可以从档案馆、文件利用中心、政府网站等基本政府信息提供单位收集用户的利用信息情况,并把这些单位的用户利用信息范围统计结果进行汇总,确定哪些政务信息是需求最强烈的,哪些是一般需求,哪些是需求很少等。其次,对政务信息资源的用户情况

进行统计分析。主要是划分用户的类型,以及掌握用户的偏好等。通过对用户的这些基本信息进行统计分析,便于从个性化和专业性等角度来构建主题地图,拓展其适应性。还要做好长期的统计工作,并根据这些统计数据预测用户的需求态势,以便构建的主题地图具有前瞻性。

(2)收集和整理电子政务信息资源。其一,对存在于不同空间的政务信息进行汇聚。按照用户对政府信息资源的需求情况,对分散在不同空间的政务信息资源进行收集。它包括物理上的收集与逻辑上的收集。物理上的收集就是把不同空间位置上的信息实体进行迁徙,并合并成实体意义上的信息集合。逻辑上的收集则是指在概念层面上的收集,它不需要迁移信息实体,但需要记录信息出处,以便能搜寻到该信息。其二,对采集的政务信息进行转换。一是对信息资源的格式、所有权、表示形式进行转换,形成可分析和可利用的政府信息资源;二是对整理好的政府信息资源进行主题内容分析,确定其主题概念,并进行安全、可靠地储存,以备构建主题地图之用。

3.3 初步生成电子政务信息资源主题地图

该阶段的任务是利用主题地图构建环节(如图2^[10]所示),初步建立电子政务资源主题地图。其核心是从准备好的电子政务信息资源中抽取主题事件,根据规范,生成初步符合主题地图规范的主题地图。

首先,在既定的数据资源规范和数据集规范的控制下,利用数据抽取环节从电子政务信息资源中抽取待处理的信息,并形成作为中间处理形态的数据集。而主题地图生成规范处理环节将这些中间形态的数据集作为“原料”,根据相应的规范和主题地图生成机制生成主题地图。电子政务信息资源主题地图的创建方式可以有两种。一种是人工方法。此种方式耗费的人力非常大,但生成的主题地图内容丰富,质量较高,较适应复杂主题地图的生成。另一种是机器自动创建。该种方式受信息源的结构影响较大。一般而言,它对结构化的信息效果较好,可以自动创建效果较好的主题地图;反之,则需要运用多种自然语言处理工具的辅助才可较好地完成^[11]。针对政务信息资源量大、来源复杂等特点,本文建议,在做好数据准备的基础上,最好采用统一标准的机器自动创建方式构建主题地图,因为它可以大大地减少工作量,缩短政务信息资源进入有效利用领域的时间,而对那些结构复杂而又不可或缺的政务信息则可以适当地采用人工方式为补充进行主题地图构建。电子政务信息资源主题地图生成后,接下来的环节是通过输出产生环节输出主题地图文件并存储。一般而言,目前构建主题地图而生成的文件多以 XTM 文件和关系数据

库文件为主,并在以后的工作中这些文件将转化为关系表的形式,并直接以数据库的形式保存^[12]。

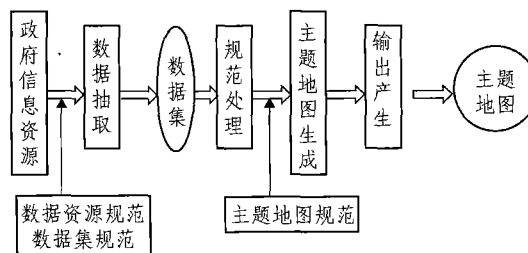


图2 主题地图初步生成模型

3.4 校验电子政务信息资源主题地图

主题地图初步生成后,接下来的工作主要是利用主题地图的验证、测试环节,根据语义验证规范,验证生成文件(如 XTM 文件)的合法性,并最终生成满足验证条件的政务信息资源主题地图。

这一步工作,可以由机器自动检验,也可以采用人工的方式来完成,并且在实际应用中这两种方式相互配合使用。其中机器检验是通过设计软件来实现,它在主题地图构建的基础上,使用主题地图的模式及约束语言(Schema and Constraint Language for Topic Maps, SCTM),自动完成对政务信息资源主题地图的语义验证,保证主题地图在语义上的合法性^[13]。这种校验方式目前应用较为普遍。而人工检验,主要是对主题地图进行实际测试,由设计者和用户来完成。设计者主要任务是测试最终形成的成果(主题地图)与最初的设计是否相一致,是不是每个主题都有意义,是不是有什么遗漏,是不是某个元素放错位置不符合语义要求等。用户主要是从使用的角度来测试,主要是看最终形成的电子政务信息资源主题地图是否便于利用,是否能帮助他们能精确地找到自己所需要的政务信息,是否比以往获得更大的收获等。通过测试、检验,找出电子政务信息资源主题地图不足的地方,并不断完善,使它真正符合实际应用。

3.5 搭建用户界面

主题地图校验工作完成之后,就要让它与用户见面,要以适当的方式表示出来。该阶段的主要工作是为用户提供基于主题地图的浏览、查询等操作的界面和接口^[14],将主题地图以包含语义的网页形式(即知识组织的形式)呈现出来,而生成的网页不但可以根据结构、语法的联系,也可以通过概念与概念之间的相互关系自由导航^[15]。这一步是电子政务信息资源主题地图面向最终利用的最后环节,也是电子政务信息资源主题地图是否实现其价值的关键。目前可以通过 Java 或 C++ 等语言工具来实现这一设计。

用户界面建设工作还要与电子政务门户网站建设有机地结合起来。要通过不断完善政府门户网站的设计,增强其智能化性能,方便用户访问。要设置电子政务信息资源主题地图利用指南,解答用户疑问,让用户更加便利地利用政务信息资源。

3.6 维护更新电子政务信息资源主题地图

日常的运行过程中,仍需要对电子政务信息资源主题地图不断进行维护更新。这是因为:第一,主题地图是一个活的东西,会随着应用领域所包含的信息(知识)的进化而进化^[16]。人们对主题的理解会不断更新,以往的某些主题可能不适合要求。这就得进行调整、更改或重新进行设计。第二,在日常运行中,电子政务信息资源系统也会出现一些问题,如数据丢失等。这些都离不开日常维护工作。第三,主题地图虽然能够较好地解决政务信息资源组织中存在的缺乏语义等问题,但在知识组织与表示方面依然存在一些缺陷^[17]。随着理论研究的深入和技术的进步,主题地图的性能必将越来越先进。这势必也要求电子政务信息资源主题地图随技术的进步而不断升级。

维护和更新工作是电子政务信息资源组织系统的一项不可或缺的重要工作,它是主题地图构建工作的延伸。通过维护、更新等措施,不断完善电子政务信息资源主题地图的性能,让它真正成为信息、知识组织和导航的有力工具,优化政务信息、知识结构,方便用户利用。

(上接第63页)题,利用 Jena 插件和 Racer、Pellet 本体推理机,在此基础上设计了 OWL 推理检索引擎,完成了军用飞机领域本体一致性检查和推理检索。

(2)解决了本体可视化递增性浏览难题,动态绘制了本体内部结构和包含的语义关系,实现本体的二维网状的图形可视化检索,动态图形浏览器支持全览、聚焦、移动等功能。

(3)给出了基于 B/S 结构和 Java 平台的本体可视化检索系统的实现方案,可将本体与现有的互联网资源整合,容易使用和共享。

(4)系统提供的关键词检索、树形列表检索、可视化图形检索等多个检索过程同步,使得系统的互动性更高,使用界面更友好方便。

参考文献

- 1 杜小勇,李曼,王大治. 语义 Web 与本体研究综述. 计算机应用,2004(10)
- 2 李景,孟连生. 构建知识本体方法体系的比较研究. 现代

参考文献

- 1,17 韩泽春. 主题地图——一种有效的知识组织与揭示方法. 忻州师范学院学报,2005(4)
- 2 Michel Biezunski, Martin Bryan, Steve Newcomb. ISO/IEC FCD 13250:1999-Topic Maps. [2006-05-23]. <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0058.htm>
- 3,7 秦铁辉等. 信息时代的“全球定位系统”——主题地图. 江西图书馆学刊,2005(1)
- 4,8 Steve Pepper. The TAO of Topic Maps. [2006-08-20]. <http://www.ontopia.net/topicmaps/materials/tao.html>
- 5 马费成,郝金星. 概念地图在知识表示和知识评价中的应用(1)——概念地图的基本内涵. 中国图书馆学报,2006(3)
- 6 Kal Ahmed and Graham Moore. An Introduction to Topic Maps. [2006-07-06]. <http://www.architecturejournal.net/2005/issue5/Jour5Intro/>
- 9,10,12,13,15 吴笑凡等. 基于主题地图的异构知识集成. 现代图书情报技术,2005(11)
- 11,16 彭志远. 基于主题地图的知识组织研究与应用[硕士论文]. 兰州大学,2004
- 14 张佩云等. 主题地图标准及其应用研究. 安徽大学学报(自然科学版),2004(3)

吕元智 上海师范大学人文与传播学院讲师,武汉大学信息管理学院2006级博士研究生。通讯地址:上海市桂林路100号。邮编200234。

王心裁 武汉大学信息资源研究中心专职研究员,教授。通讯地址:武昌珞珈山。邮编430072。

谭必勇 武汉大学信息管理学院2005级博士研究生。通讯地址同上。(来稿时间:2006-11-02)

图书情报技术,2004(7)

- 3 IEEE Working Group. Suggested Upper Merged Ontology (SUMO). [2006-03-20]. <http://www.Ontologyportal.Org/>
- 4 颜端武,丁晟春等. 基于语义 Web 和 Jena 插件的语义检索系统实验研究. 情报理论与实践,2006(3)
- 5 Pellet OWL Reasoner. [2006-03-20]. <http://www.mindswap.org/2003/pellet>
- 6 SPARQL Query Language for RDF. [2006-04-10]. <http://www.w3.org/TR/rdf-sparql-query>
- 7 TouchGraph Development. [2006-03-10]. <http://touchgraph.sourceforge.net>
- 8 Ajax: A New Approach to Web Applications. [2006-02-08]. <http://adaptivepath.com/publications/essays/archives/000385.php>

颜端武 岑咏华 在读博士,南京理工大学经济管理学院信息管理系讲师。通讯地址:南京。邮编210094。

毛平成 晓 南京理工大学经管学院信息管理系情报学硕士生。通讯地址同上。(来稿时间:2006-11-15)