

●黄如花

## 网络信息组织的发展趋势<sup>\*</sup>

**摘 要** 理想的网络信息组织模式应该以用户为中心,遵循实用性和易用性原则,综合运用自然语言和人工语言,充分利用新兴技术和人们经验的积累,朝着信息描述标准化、组织活动合作化、组织技术智能化、资源与服务集成化、用户界面可视化、内容揭示深入化、组织方法多样化方向发展。图1。参考文献30。

**关键词** 信息组织 知识组织 书目控制 因特网 发展趋势

**分类号** G350

**ABSTRACT** The author thinks that an ideal network information organization pattern should include user-centeredness, practicality, combination of natural and artificial languages, full use of new technologies and human experiences, and development towards the standardization of information description, the cooperation of organization activities, intelligence of organization technologies, the integration of resources and services, the visualization of user interfaces, the deepening of content description, and the diversity of organization methods. 1 fig. 30 refs.

**KEY WORDS** Information organization. Knowledge organization. Bibliographical control. Internet. Development trends.

**CLASS NUMBER** G350

20世纪90年代以来,图书情报界、计算机领域的专家和从业者共同努力,使得网络信息组织的方式与方法产生了许多变化:信息组织的对象从各种类型的数据发展到具有丰富内容的知识,组织形式从数据结构发展到知识表示,组织方式从单一发展到多样化,组织的结果从静态的文本格式发展到动态的多模式的链接。网络信息组织方法体系已经初步形成,但网络信息组织方式与方法仍没有圆满解决网络信息的序化与优化问题。如何使纷繁复杂、无序的网络信息资源逐步向有序化、便于用户索取的理想状态发展,是一个亟待解决的现实问题,也是现代信息资源管理的重要内容。那么,未来的网络信息组织路向应该如何呢?

理想的网络信息组织模式应该以用户为中心,遵循实用性和易用性原则,综合运用自然语言和人工语言(分类语言、主题语言),充分利用新兴技术和人们经验的积累,朝着更符合用户需要的方向发展。具体来讲,应该朝着以下7个方向发展。

### 1 信息描述标准化

网络信息具有数量庞大、质量参差不齐、媒体与格式多样、容易传播与共享等特点,这就要求我们必须遵守统一的信息揭示规则,以促进网络信息资源的充分利用和用户信息需求的满足。网络信息组织主要方法中的文件、搜索引擎、编目、学科信息门户均涉及标准化问题。标准的制订与推行是解决这些问题的惟一出路,也是网络资源共享的必要前提。

#### 1.1 图书情报机构的参与

国际图联制定了《IFLA书目记录的功能要求》(IFLA's Functional Requirements for Bibliographic Records)。OCLC一直注重研究、宣传与推行书目活动标准。国际知识组织协会(ISKO)制定了知识与信息组织的相关标准,并在其主页发布<sup>[1]</sup>。美国国会图书馆于2002年5月提出了“元数据输入与传输标准”(Metadata Encoding and Transmission Standard, METS)<sup>[2]</sup>,并设立Z39.50维护机构<sup>[3]</sup>和MARC办公室<sup>[4]</sup>。美国图书馆协会的兴趣与活动之一是标准化与指导,并有专门网页提供标准化信息<sup>[5]</sup>。美国的国家数字图书馆联盟(NDLF)、图书馆及信息技术协会(LITA)、网络化信息联盟(NIF)在推进网络信

<sup>\*</sup>本文为教育部人文社会科学“十五”重大项目(2000ZDXM870001)论文之一。

息组织标准化方面也卓有成效。我国在文化部的召集下,本着统一的规划、统一的技术标准及统一的运行规则等原则,组建了“中国数字图书馆工程建设联席会议”来协调工程的资源建设和标准规范。我国已经在网络信息资源管理标准化方面迈出了重要的一步。

## 1.2 国际和各国标准机构的推动

SC9 是国际标准化组织的 ISO TC 46 的分委员会,它负责发展和维护关于文献展示、识别与描述的国际标准,并经常在其网站公布关于电子文献(包括网络信息资源)书目控制各方面的 ISO 国际标准草案,已经制定的与信息组织相关的标准有: MARC 格式(ISO2709)、SGML 格式(ISO8879)、数据要素规范与标准(ISO11179)、语言名称表示代码(ISO639)和国家名称表示代码(ISO3166)等<sup>[6]</sup>。ISO 还成立了元数据工作组,负责元数据的标准与规范工作。

美国全国标准化协会(ANSI)下设全国信息技术标准委员会(National Committee on Information Technology Standards),从事有关元数据的命名、标识、定义、分类和注册等工作,还成立了信息基础设施标准座谈小组(Information Infrastructure Standards Panel)<sup>[7]</sup>。欧盟和英国的相关机构有信息社会标准化系统(ISSS)和英国标准协会的向用户传递信息解决方法部。

另外,因特网管理机构、不同学科领域的学(协)会、公司、民间自发组织甚至个人,都在为网络信息组织及其相关的信息交换、信息检索、通信协议等方面标准的制定与推行作出积极努力。万维网联盟(W3C)是万维网上最有影响的因特网标准的认定机构,在网络信息组织领域,该机构认可的网络资源描述语言为 XML,资源描述框架为 RDF,元数据标准为都柏林核心元数据(DC),日期与时间格式为 W3CDTF<sup>[8]</sup>。国际信息与图像管理协会(Association for Information and Image Management International, AIIM)<sup>[9]</sup>和数据交换标准协会(Data Interchange Standards Association)<sup>[10]</sup>则分别负责图像信息、信息交换相关标准的制定。其他的因特网管理机构还有:因特网协会<sup>[11]</sup>、因特网网络化信息中心<sup>[12]</sup>和因特网结构署<sup>[13]</sup>等。

因特网工程任务组(IETF)主要负责有关因特网的各种技术标准及接口规范的界定,所发布的各类标准与协议以 RFC(The Request for Comments,即征求意见稿)定名,表明其作为民间机构而不是官方标准界定机构的地位,但事实上大多数 RFC 都已成为网络界的事实标准。它发布的与网络信息组织相关的标准有:用于语言标识的头标(Internet RFC1766)、统一资

源标识符——一般语法(Internet RFC2396)和用于资源发现的都柏林核心元数据(Internet RFC2413)。该机构中与网络信息组织相关的工作组有:目录的存取、检索与索引工作组,通用索引协议工作组,电子数据交换——因特网整合工作组,因特网信息资源整合工作组和统一资源识别符工作组等,2001 年还召开国际会议探讨确立网络实名标准<sup>[14]</sup>。微软及几个大型的电子书生产商于 1998 年联合提出了界定电子书内容格式标准的 Open eBook 计划<sup>[15]</sup>。

国内外在网络信息组织标准化方面已取得一些进展,但仍有必要形成一系列标准与准则,使网络信息组织活动有规则可依。

## 2 组织活动合作化

因特网是一个多网络、无主管的分散型互联结构,网络信息的通畅流动与有效利用要求各方面的整体配合。网络信息的组织是一项涉及面广的持久性工作,需要世界范围内的合作以保证准确、及时地报道网上信息及其变动情况、提高信息的质量并实现规模效益。网络信息组织的合作将得到强化。

图书情报界已经开展了网络信息组织合作化的有关活动。国际图联发起了“全球书目控制和国际机读目录核心活动(Universal Bibliographic Control and International MARC Core Activity, UBCIM),并编辑出版了《国际编目与书目控制》(International Cataloguing and Bibliographic Control, ICBC)杂志。2002 年 7 月在西班牙的 Granada 召开的第 7 届国际知识组织大会的主题是“21 世纪知识表示与组织的挑战:跨越边界的知识一体化”<sup>[16]</sup>。OCLC 一直在书目控制与资源共享的协作中扮演着国际中心的角色,国际图书馆协作体联盟(International Coalition of Library Consortia, ICOLC)则是将各种协作体组织起来的组织机构<sup>[17]</sup>。

网络信息组织的合作还将超出图书情报界,扩大到整个信息生产链上的其他所有参与者,包括网站内容创作者、出版商和信息资源系统的合作和集成者等。美国的网络图书馆(NetLibrary)是世界上最大的全文电子书的收藏者与服务提供者,也是出版商、发行商、图书馆与读者有效结合的典范<sup>[18]</sup>。BIBLINK(出版机构与国家书目服务连接),是由英、荷、挪、法、西班牙 5 个国家图书馆等 9 个机构合作的项目。该项目解决了出版机构书目数据与 MARC 数据的双向转换,有利于实现将出版机构提供的数据作为电子出版物网络元数据的构想。

### 3 组织技术智能化

传统的信息组织大多是以人工方式进行的,其中著录、标引以及分类表、词表的编制和维护等都是烦琐的手工劳动。网络信息的特点不允许过多的中间和加工环节,而要求在信息组织的诸环节利用自动化技术。随着网上自动分类、自动标引、自动编制分类表与词表、自动漫游技术,信息类别的自动判别技术和信息推拉技术的逐步发展完善,会有越来越多的网络信息资源被自动地追加、组织到相应的位置,方便用户及时准确地检索信息。例如,目前已经出现了自动抽取或产生元数据的软件:对元数据的自动抽取与收集(Automatic Extraction/Gathering of Metadata)<sup>[19]</sup>,和元数据的自动产生(Automatic Production of Metadata)<sup>[20]</sup>。人工智能技术与这些技术的结合,使得用户的信息需求可以被智能地推理与理解,促进网络信息组织智能化的发展与应用。

一种智能化的书签软件(PowerBookmarks)已经面世,它可以从作为书签的资源URL中解析出元数据,并自动对该资源进行标引与分类。书签软件系统通过监测和利用用户的信息搜索与浏览过程,自动获得用户的需求与兴趣信息,把合适的信息提交给用户,并允许用户订购感兴趣的新资源或更新已有的信息资源。智能浏览器、智能代理等智能产品的出现创造了一种智能合成环境(Intelligent Synthesis Environment, ISE),ISE是新一代分布式信息系统的代表,它使用成组的虚拟现实交互设备提供沉浸式的接口,提供一个多媒体环境和对信息源的统一个性化访问,以人们习惯的方式支持信息的利用。

### 4 资源与服务集成化

网络资源的多样性和异地性,影响了用户对网络资源的有效利用,各种网络资源整合于同一个界面将成为发展的必然。海量信息集成化系统通过一致的对外接口,使用户能方便快捷地浏览和访问各种异构信息,比如文档信息、电子邮件、共享代码、多媒体信息和数据库信息,从而节省用户访问分布于不同地点资源的时间和精力,网络信息资源将和其他各种信息资源一起被组织成集成化的信息系统,便于用户在一个资源丰富的“信息超市”(information supermarket)浏览和选择自己所需信息,并向用户提供一站式检索(one-site search),如图1所示。

用户的信息需求已从单纯的文本信息提供向交

互式多媒体信息发展,从书目信息的检索、全文的查看、全文的传递等分别服务向文本(含全文)、语音、图像集成化的服务而一站传输为主,这就要求组织后的信息系统便于用户在同一界面上获得多种服务。实际上,一些集成化的试验或者应用系统已经出现,比较突出的如美国的OhioLink<sup>[21]</sup>、英国的National Electronic SiceLicense Initiative<sup>[22]</sup>和Distributed National Electronic Resources(DNER)<sup>[23]</sup>以及我国的CALIS等。

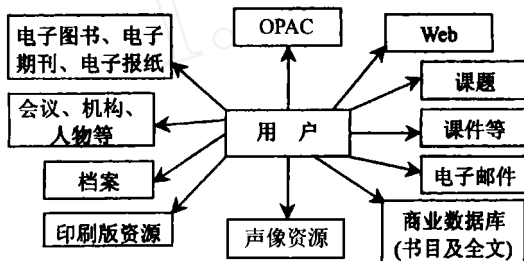


图1 集成化的信息组织与服务

杭州麦达数据公司推出的数字化平台可将万方数据资源系统、维普公司数据库与图书馆已有资源整合于一个统一的界面,采用该平台的各单位之间还可实现对他馆资源的无缝访问<sup>[24]</sup>。

### 5 用户界面可视化

可视化是信息组织的透明化与易用性所要求的。由于多数用户缺乏计算机数据处理知识和必要的信息检索技能,他们要求信息组织方式和提供的检索简便易行。

网络信息组织中用于指明信息资源方位与联系、具有导航功能的信息地图便是一种可视化工具,它对于信息迷航的解决优于其他方式。相对于搜索引擎,信息地图不仅可以揭示信息的方位,而且能够体现它们之间的联系(在某些情况下,这些联系更具重要性),还能帮助使用者从全局上了解信息的分布状况。信息地图的具体表现形式多种多样,如有向图、树型图、扇型图等。

随着基于Windows的操作系统、按钮和菜单的用户界面的引入,用户与计算机之间的交互也发生了改变。2001年10月问世的Windows XP比此前的Windows界面更具图形化,提供了更为自然的人机界面,使得让机器适合人而不是相反成为可能。如果说Windows为可视化信息组织提供了技术环境,那么,JAVA、Visual B、Visual C+、HTML和可视化标准建模语言(Unified Modeling Language, UML)等语言的使用

以及动画技术、三维技术和虚拟现实技术的产生和发展,则有利于推动可视化信息组织的快速发展。

目前,可视化信息组织技术已经在地理信息系统、产品设计、数字城市、数字地球等领域得到广泛应用,有些图书馆或档案馆的查询界面也运用了可视化技术。

## 6 内容揭示深入化

数字时代用户对信息检索的需求不再仅仅满足题名、作者、主题词等传统条件下有限的检索点,而更注重实际内容(如目次、提要、文摘、全文、知识点等)的检索。这就对信息组织提出了更深入化的要求。

### 6.1 组织的对象深入化

迅速普及的网络技术和数字技术使任意层次的任意信息元素、信息单元和信息集合体系正在逐步以计算机可识别和可理解的方式被定义、描述、指向、链接、传递和动态组织。网络信息组织的对象不仅停留在对信息特征的描述,而且深入到知识单元,扩大标引广度,增加数据库的标引深度,通过多层次、多方位的描述与分析来揭示与组织网络信息资源,以促进网络信息资源的合理利用。

### 6.2 网络信息挖掘

传统的数据挖掘技术所涉及的主要是结构化的数据库,而目前网络上有 90% 的可用信息是非结构化信息,且网上数据往往是经常变动和不规则的。网络信息挖掘在已知数据样本的基础上,通过归纳学习、机器学习、统计分析等方法得到数据对象间的内在特性,据此采用信息过滤技术在网络中提取用户感兴趣的信息,或者更高层次的知识和规律。它除了处理传统数据库中的数值型的结构化数据外,处理更多的是文本、图形、图像、www 信息资源等半结构、非结构的数据。Google 便是采用网络挖掘技术判断网站的重要程度并据此排列检索结果的。

人工智能技术和信息推送技术促进了网络信息的挖掘与深层次揭示,以更好地满足不同用户的各种需求,如系统自动运行,不断更新用户的资料库,提供个性化的主动信息服务。从信息中采掘知识,再将知识变成社会财富,应该成为网络信息组织与服务的新方向。

### 6.3 向知识组织发展

未来的网络信息组织要更严格地控制信息的质量,对网上信息进行有效评价和筛选,为用户提供有价值的信息,而不是大量的无用的信息,其目的是向人们提供便于利用的、可以帮助解决问题的序化的知识,实现从信息层次到知识层次的根本转变,组织

的知识包括显性知识与隐性知识。目前的知识组织主要以文献单元为基础和以数据(各种事实、概念、数值的总和等)单元为基础,但都是静态的、列举式的。未来的知识组织将以专家系统为基础,具有动态联系、判断、分析、比较、推理等新型的知识处理与组织功能。美国的数字图书馆二期工程已开始创建“网络化的知识组织系统与服务(Networked Knowledge Organization Systems and Services)。

向知识组织发展的另一个新动向便是运用知识管理的思想,以“知识库”(knowledge base)或“知识门户”(knowledge portal)取代现有的“信息结构”(information architecture),而建立起电子知识空间(e-knowledge)。较成功的范例有俄亥俄州立大学建立的“知识库(Knowledge Bank)<sup>[25]</sup>和 MIT 的“数字空间项目”(Dspace Project)<sup>[26]</sup>等,这样的成果还有助于推进远程教育 with 终身学习。

## 7 组织方法多样化

### 7.1 面向内容的信息组织

这是目前用得最多的方法,大多数网站和搜索引擎都采用这种方法。它专注于有关面向学科(专题)的资源如何组织与发现。其中,相对于大型的综合性网站而言,对某专题信息进行集中而深入揭示的学术信息门户(scholarly portal)将受到更多关注,其发展也会更快。

### 7.2 面向对象的信息组织

面向对象方法“就是以对象为中心、为出发点的方法”,“对象就是我们在问题空间中要考虑的那些人或事物。”<sup>[27]</sup>网络环境下,用户需求变化除了有需求量上的增长外,还表现为信息需求复杂性的提高。同一个用户在学习、娱乐、工作等不同的活动中也有着不同的信息需求,希望有一个系统能直接、深入、有效地支持其检索、处理和利用信息来解决问题,帮助建立个人的数字图书馆。

事实上,基于网络的个性化信息组织与服务的思想在国外网站设计或信息服务系统中已经有一定的发展。雅虎推出了面向儿童的搜索引擎 Yahoo!ligans。许多门户网站和信息提供者推出了个性化定制服务系统,如 My Yahoo!, My Lycos! 等。国内许多网站为了满足国内读者和英语国家用户的不同需要,将其主页以多种版本组织(简体中文、繁体中文与英文版),这些做法体现了以用户为中心的思想。面向不同对象的信息组织也受到了图书馆界的重视,他们开发出了 My Gateway, My Library 等个性化

定制服务系统,如美国北卡罗来那州里大学(NCSU)图书馆的 MyLibrary<sup>[28]</sup>,麻省理工学院和中国科学院都已建立起基于用户的数字图书馆。

随着 Web 数据库技术、信息推送技术和智能代理技术等个性化信息服务所需的支撑技术的逐步成熟,网络信息组织将发展成一种信息代理服务,即根据网络信息的属性和用户需求,对网络信息进行加工、整理、排列、组合,使之有序化,以满足用户对网络信息的需求。

### 7.3 面向任务的信息组织

即面向特定用途的信息组织。企业信息门户(Enterprise Information Portal, EIP)将企业内部的信息系统与外部信息系统有机整合,将贸易、价格、法规、税收、产品、市场、管理、竞争对手等多方面信息汇集于一身,服务于企业的经营管理任务,并通过对事件和信息处理,把用户有机地联系在一起。我国各地的信息港或地方热线则汇集当地的历史、地理、旅游、人文、物产、政治、风土人情等方面的信息。基于网络的电子政务信息系统则是从政府角度出发的面向政府机关、企业以及社会公众的信息服务和信息处理系统,包括政府的信息服务、电子贸易、电子化政府、政府部门重构、群众参与政府多个方面的内容。全球最大的政府网站——美国的 First Gov 按照美国电子政务的功能来组织信息:一是用于政务公开的信息;二是用于网上服务的信息,细分为市民的服务、为企业的服务和为政府的服务 3 个板块;三是用于资源共享的信息;四是政府内部办公电子化的信息;五是提供安全保障的信息<sup>[29]</sup>。

未来的网络信息组织模式还将朝着管理分布化等方向发展。当网络成为更为成熟、健康的信息生产与使用的信息空间时,网络信息资源的品质将得以提高,信息组织者在信息筛选与过滤上的劳动将会减少。事实上,人们已经在为网络环境的改善作出实际行动。我国先后颁布了一系列与互联网信息生产与服务有关的法规,如:《互联网信息服务管理办法》、《互联网电子公告服务管理规定》和《互联网站从事登载新闻业务管理暂行规定》等。为了规范我国互联网行业从业者行为,中国互联网协会于 2002 年 3 月制定了《中国互联网行业自律公约》<sup>[30]</sup>。新闻出版总署和信息产业部颁布的《互联网出版管理暂行规定》已于 2002 年 8 月起正式实施。

### 参考文献

1 <http://is.gseis.ucla.edu/orgs/isko/stds.html> (accessed

Dec. 1, 2002)

2 <http://www.loc.gov/standards/mets/> (accessed Dec. 16, 2002)

3 <http://lcweb.loc.gov/z3950/agency/> (accessed Dec. 18, 2002)

4 <http://lcweb.loc.gov/marc/marc.html> (accessed Dec. 18, 2002)

5 <http://www.ala.org/work/standards.html> (accessed Dec. 18, 2002)

6 <http://www.nlc-bnc.ca/iso/tc46sc9/index.htm> (accessed Dec. 18, 2002)

7 <http://web.ansi.org/public/iisp/> (accessed Dec. 18, 2002)

8 <http://www.w3c.org/> (accessed Dec. 16, 2002)

9 <http://www.aiim.org/> (accessed Dec. 16, 2002)

10 <http://www.disa.org/> (accessed Dec. 16, 2002)

11 <http://info.isoc.org/> (accessed Dec. 16, 2002)

12 <http://www.internic.net/> (accessed Dec. 16, 2002)

13 <http://www.iab.org/iab/> (accessed Dec. 16, 2002)

14 <http://www.ietf.org> (accessed Dec. 16, 2002)

15 <http://www.openebook.org> (accessed Dec. 10, 2002)

16 <http://is.gseis.ucla.edu/orgs/isko/news.html> (accessed Dec. 18, 2002)

17 <http://www.library.yale.edu/consortia/icolcpr.htm> (accessed Dec. 18, 2002)

18 <http://www.netlibrary.com> (accessed Dec. 18, 2002)

19 <http://www.ukoln.ac.uk/metadata/dcdot/> (accessed Dec. 18, 2002)

20 <http://www.klarity.com.au/> (accessed Dec. 10, 2002)

21 <http://www.ohiolink.edu> (accessed Dec. 18, 2002)

22 <http://www.nesli.ac.uk/> (accessed Dec. 18, 2002)

23 Adding Value to the UK's Learning, Teaching and Research Resources: the Distributed National Electronic Resources (DNER). <http://www.jisc.ac.uk/pub99/dnervision.html> (accessed Dec. 12, 2002)

24 <http://www.metadata.com.cn> (accessed Dec. 10, 2002)

25 The OSU Knowledge Bank. <http://www.lib.ohio-state.edu/KBinfo/> (accessed Dec. 18, 2002)

26 Dspace Project. <http://www.dspace.org/live> (accessed Dec. 18, 2002)

27 王汉新,王道平.管理信息系统教程.北京:中国劳动社会保障出版社,2000,148~150

28 <http://my.lib.ncsu.edu/> (accessed Dec. 13, 2002)

29 <http://www.firstgov.gov> (accessed Dec. 18, 2002)

30 中国互联网行业自律公约.光明日报(电脑网络周刊), 2002-03-27

黄如花 武汉大学信息管理学院副教授、博士。通讯地址:武汉市珞珈山。邮编 430072。

(来稿时间:2002-12-30)