

潘燕桃

The Scout Report Signpost 因特网编目分析研究

摘要 The Scout Report Signpost(SRS)是利用现有的分类主题工具,采用题名、主要URL、评述日期、责任说明等13个项目对因特网资源进行编目的成功例证。举例分析了13个著录项目,指出目前仍然存在缺乏标准权威的编目规则,难以确定日期信息等问题。参考文献12。

关键词 Scout Report Signpost 因特网编目 因特网资源

分类号 G254.3

ABSTRACT The Scout Report Signpost is a successful example of the cataloging of Internet resources. In this paper, the author analyzes thirteen descriptive elements, including title, main URL, review date, responsibility, etc., and discusses some problems, such as the lack of standard cataloging rules and the difficulty to determine date information. 12 refs.

KEY WORDS Scout Report Signpost Internet cataloging Internet resources

CLASS NUMBER G254.3

Scout Report Signpost(可译作侦察报告路标,以下简称SRS)是Internet Scout Project(以下简称ISP)提供的一项信息服务。ISP是一个由美国国家自然科学基金会资助,本部设在威斯康星大学麦迪逊分校计算机科学系大楼内,为大专院校的师生和科研人员筛选过滤因特网上最新学术研究和教育资源的科研项目,提供侦察报告、侦察报告路标和侦察工具箱等11项信息服务。

作为一个新的因特网资源发现的工具,SRS首发在1997年6月,主要为指导美国高等教育工作者使用高质量的电子资源,同时它也是特意为方便使用过去4年侦察报告(Scout Report,以下简称SR)的档案而设计的。SR是ISP提供的一份汇集因特网最新消息的电子周刊,图书馆员与主题专家根据一定的资源选择标准,为研究和教育群体筛选和评述最新和新近发现的因特网资源,并以环球网、gopher、远程登录、FTP、新闻组和电子邮件讨论清单等形式发送出去。自1994年4月首发至今,SR以其敏锐的触角、独到的评述、高质量的出品,为大专院校的师生和科研人员节省了大量宝贵的时间,成为他们教学、学习和科研及时而得力的助手,被誉为因特网上的资格最老、而且是最获好评的网上出版物之一。

设计SRS的另一个深层目的是开发一个能够

与其他数据库配合动作、互用的数据库。这将促进与其他信息提供构成之间的未来的合作关系。SRS是ISP精心培养出来的“尖兵”,它集中了因特网上学术资源最精华的部分,对引导美国高等教育成为高质量的电子信息资源起着非常积极的作用。

1 SRS的资源访问方式

访问SRS的资源有4种基本方式:快速检索、LCSH主题词浏览和LCC分类号浏览、高级检索。

1.1 快速检索

通过快速检索,用户能够检索到每条最低限度记录中的题名、主要URL、侦察报告摘要、评述日期和最后一次查实URL的日期等5个数据项目。快速检索通过以Excite为网络服务器的搜索引擎,可对迄今为止所有ISP出版物评述过的、包括最低限度和完全编目的因特网资源进行全文与自然语言检索。由于被评述资源中的专业术语经常被用在每一份侦察报告摘要之中,检索者可以用熟悉的、具体某一个学科的、或者其他非控制词汇进行检索。Excite索引每周编制1次,定期把1周以来最新的ISP出版物的内容以及其他已校正的URL汇总。SRS在快速检索的简要描述的题名项中提供一个超文本链接到该资源。

1.2 浏览方式

浏览方式包括LCSH 主题词浏览和LCC 分类表浏览。从SRS的主页,用户可以在LCSH 主题词浏览选择A—Z中的一个链接来浏览侦察报告摘要。这部分提供链接通往按字母顺序排列的、根据资源题名标引的LCSH 主题词的目录。资源题名最多会被标引5个主题词。

LCC 分类表浏览,即根据LCC 分类表组织起来的主题类目表进行浏览。SRS 中的资源最多会有两个分类码,每一个分类码由1~3个字母组成。这些分类码按照LCC 分类表中的字母顺序排列。需要说明的是,资源没有完整的索书号,是因为这两种访问方法只提供浏览,资源只是以简要描述的形式显示。

这两种访问方式的特别之处在于,它们提供选项直接访问资源本身和资源在SRS 中的完全描述屏幕。

1.3 高级检索

高级检索是关于以下9个关键项目(按顺序排列)的字段检索:题名、责任者—投稿者、LCSH 主题词、资源类型、语言、资源定位、出版者以及主要URL。高级检索的结果显示为SRS 记录的一个简要描述格式,这个格式仅列举题名、主要URL、最多不超过40个字的侦察报告摘要原文等数据,以及一个通往该资源的、包括侦察报告摘要全文和其他12个项目的完全描述的连接。

2 因特网编目现行标准研究

编目成果包括使用标准工具来描述和访问,这包括《英美编目条例》(AACR2R)、LCSH、《主题编目手册:主题词表》,以及LCC 分类表来确定款目和名称权威的格式,著录适当的主题词和分类号。由于MARC 格式不再用作数据库结构,SRS 设计了一个假设变换表(hypothetical mapping)将SRS 项目变换成为都柏林核心项目和具体的MARC 字段。一个SRS 项目有时能够变换成几个MARC 字段。在假设变换表列举了最具有意义的、在确定每一个SRS 项目的款目组成需要时候考虑的语法标准。然而,除了赋予电子资源性质和SRS 非MARC 数据库结构,这个变换表仅是在理论层面的,还未被应用在所有方面。

AACR2R 把编目条例分成两部分:第一部分用

于说明款目的形式;第二部分是检索点的确定及其款目的形式。每一部分的条例都适合于不同的MARC 字段。而且,用来说明一个MARC 字段的数据通常都需要从来源逐一输入。然而,对于一个访问字段,款目的形式则是来自一个权威档或一个详尽条例中的控制语句结构。因为SRS 没有利用MARC 数据库结构,也没有把它的字段类型分成说明行和访问行,有时候需要把MARC 中的两个款目形式换成一个SRS 的项目。例如:在一个MARC 格式的数据库里,“尼尔森·曼德拉就职演说”网站的责任说明信息是“尼尔森·曼德拉”;应把它输入说明MARC 字段(245 \$c)。考虑到主款目(Main Entry)(100)会按照“曼德拉, 尼尔森, 1918—”输入。SRS 规定所有的著录项目都必须是最完整的项目。因此,在这个例子里,SRS 款目的格式即是“曼德拉, 尼尔森, 1918—”。SRS 的这个编目原则能够帮助用户更容易地访问到最完整的信息。

3 SRS 中的因特网编目结构分析

作为一个因特网资源编目成功的例证,设计SRS 的目的之一便是测试传统的编目条例如何能够适用于因特网资源,如何与新出现的标准匹配。ISP 在参考了都柏林核心的有关规定之后,决定设计非MARC 数据库并使用著录项目来进行编目,同时作为检索SRS 的入口。ISP 分别对不同的因特网资源进行最低限度编目和完全编目。

3.1 最低限度编目(Minimal Cataloging)

SRS 对被评述过的侦察报告、网络动态、儿童报告和侦察工具箱4种ISP 出版物中的所有资源进行最低限度编目。从1996年12月开始,侦察报告内容的编目成果正式在网上出版,当时共有1900个资源。现在这个文档的编目仍在继续,但每个资源只有最低限度记录。这些最低限度记录的数据显示在简要描述屏幕上,并且可从快速检索的检索结果得到。这些最低限度编目记录和SRS 简要描述包括5个数据项目:题名、主要URL、侦察报告摘要、评述日期以及最后一次核实URL 的日期。

第一题名显示符合SRS 的编目标准,也许跟原侦察报告摘要的大题名一致或不一致。最低限度记录中的题名、主要URL 与侦察报告摘要等项目来自ISP 出版物中的原始摘要。原侦察报告摘要由标

目以及随后的每一个资源评述的说明文字组成。每一条书目都包括被评述资源的题名、辅助应用软件或即插即用软件、括号中相关的资源,以及有关的 URL。

3.2 完全编目(Fully Cataloging)

第一代 SRS 编目数据库应用 Claris Filemaker-Pro 制作,截止 1999 年 5 月 6 日共有 3335 个完全编目的因特网资源。SRS 中所有的数据都是原始编目,完全编目的资源每周平均递增 50 个。为了对因特网资源进行完全编目,SRS 采用了题名、侦察报告摘要、主要 URL 等 13 个基本项目或字段。这些项目基本上来自都柏林核心元数据项目集(Dublin Core Metadata Element Set),根据实际需要,SRS 将某些项目做了一些相应的改动。下面举例逐一分析这些项目。

3.2.1 题名

SRS 完全描述屏幕中的题名项目含有一个直接连通资源的超文本链接。为了最大限度地增加可视性,SRS 没有遵循 AACR2R 中 A.4A.1“仅题名的第一个字母需要大写”的规定。如果资源的题名本身没有大写,SRS 就会原样输入。

题名有时会不同于原侦察报告摘要,这主要有 3 种情况:(1)在第一次被评述到被编目期间,资源的题名被改动了;(2)有些资源常常有几个题名,侦察报告的编辑人员选择了其中一个,而这个正好异于原侦察报告摘要中的题名;(3)为了使题名一目了然,SRS 有时会将题名合成。当一个资源有多个题名时,SRS 通常会选用最完整的题名作为题名项目,同时把其他的题名信息输入到互见题名项目。

这个互见题名并没有在网页上显示。但是,当题名检索在高级检索中执行时,互见题名项目的内容就会与题名项目的内容一起显示出来。如果一个组织或机构在资源题名中被提及,而公众都知道它的首字母缩语,这个信息就会被输入互见题名项目。相反地,如果题名只是个缩语,它就会在互见题名字段扩充成全称。设立这个项目的目的是为了使用户在检索时,不会因为知道的只是全称或是缩语而受到限制。

编目人员应该考虑一个环境网资源至少有 4 个不同题名的可能性: title 标记(显示在浏览器窗口的题名栏)、图形网页的题名、用于题名图形的 alt 标记,以及用于图形网页的纯文本版本的题名。

3.2.2 侦察报告摘要

除了其他必需的辅助应用软件或即插即用软件外,资源题名依照侦察报告摘要的描述部分,URL 都会一一列出。

一份侦察报告摘要出现多于一个 URL 的原因主要有(1)一个资源同时有图形版本和纯文本版本的 URL。这主要是为了满足 ISP 读者不同的访问需要。(2)一份摘要也可能描述两个或更多主题相关的资源,有时这些资源也会被另外编目,同时附带每一条数据库记录显示原始摘要全文和所有相关的 URL。(3)一份摘要可能着重强调一个资源,这个资源属于一个巨大资源的一部分,这样,摘要中就会常常包括被摘要的资源 and “母体”资源的 URL。

3.2.3 主要 URL

一个主要 URL 可以指示环球网、gopher、FTP 和远程登录的站点,以及一份电子邮件清单和一个新闻组的地址。当有几个 URL 出现在侦察报告摘要的时候,只有一个是被指定的主要 URL。有些资源能够通过多于一个的网络协议来进行访问。如果一个资源能够同时在一个环球网站点和 gopher 访问,那么环球网站点的 URL 就会优先成为主要 URL 项目的款目。其他所有的网络协议仍会出现在 SRS 完全描述屏幕的侦察报告摘要中。

ISP 期望不久的将来能够出现一个更为固定的解决模式,如统一资源名称。一旦有类似的标准广泛推广实施,ISP 将会适当修改这个项目的名称及内容。

3.2.4 评述日期

这个项目指示资源出现在 ISP 出版物的日期。该项目含有一个超文本链接,从 SRS 的完全描述屏幕直接连通 ISP 出版物的档案。为了容易理解,这个日期在 SRS 中被标记为月/日/年(MM/DD/YY)。最近,都柏林核心工作小组决定使用“年/月/日”(YYYY/MM/DD)的格式来标记这个项目。

3.2.5 最后一次核实 URL 的日期

这个项目表明最后一次通过自动方式或手工方式核实主要 URL 的日期。这是自动程序检查网站的主要 URL 项目的数据,但它并不能够立即查明使用不同协议的所有资源的日期,如电子邮件清单,被输进这个项目的日期只是这个资源原来出现在相应的 ISP 出版物上的日期。最近,DCWG 决定使用“年/月/日(YYYY/MM/DD)”的格式来著录这个项目。

3.2.6 资源类型

有分析表明, SRS 的内容揭示了因特网资源的两种主要特征: (1) 格式, 包括文档、数据库、常问问题等; (2) 信息内容, 由教育类资料、书目信息、会议—征求意见等组成。SRS 把因特网资源大致分为 17 种组合资源类型: 有不少资源同时属于一种以上的资源类型, 没有上限规定。然而, 为了分类准确, 必须核实格式和内容。例如: 一个网站中虽然有一部分注明是“FAQ”, 但这方面的内容只占很小比重, 那么该网站就不会被自动归到资源类型的“FAQ 类”。17 种资源类型如下:

动画—录像: 动画文件可能需要辅助应用软件或即插即用软件;

音频产品: 声音文件可能需要辅助应用软件或即插即用软件;

目录: 印刷版和(或)电子版源文的引文目录;

图表: 图形或列表数据, 主要的统计数据 and (或) 图表;

会议—征求意见: 会议消息、出版公报、计划征集、经费申请等;

数据库: 网站、gopher、远程登录站点等中的能够检索的信息集;

辞典—百科全书: 带有定义或解释文字的词汇和术语表;

电话号码簿: 个人、机构等的联系信息目录;

文件: 报告、文章、评论等;

10 教育类资料: 课程建设、课程资料等;

11 常问问题 (FAQ): 经常被问及的问题;

12 图形: 图画、照片、影像等;

13 杂志—报纸: 期刊(纯电子期刊和印刷期刊的电子版)、期刊文章汇集、目录、文章索引;

14 图书馆目录: 图书馆资料的交互式电子目录;

15 邮递清单—新闻组: 利用电子邮件的讨论小组、电子公告栏等;

16 元站点(Meta-site): 已被评述或还未被评述的因特网上资源的指南;

17 软件: 能够被下载的应用软件、脚本(Scripts)等。

用户能够从任何一个 SRS 完全描述屏幕直接连通已被编目的、在同一类目下的所有记录。寻找具体某一方面的资源类型种类的用户也能够的高级检

索中限定一个具体类型来进行检索。

3.2.7 责任者

因特网资源的责任者信息常常是含糊不清或不完整的。如果没有明确的关于资源信息内容的责任者说明的话, 根据 AACR2R 关于主要款目的条例, SRS 就会使用“没有有关信息”的语句。

资源中的“由...制作”或“由...设计”等语句并不能说明他们就是信息内容的责任者。例如: 当“由...制作”中间是一个名称的话, 有可能是指责任者, 但也有可能是指文件格式(如 HTML 或 PDF 等), 或是资助来源。在上述的责任者不明确的情况下, 这个信息就会被输入投稿者项目。然而, 需要说明的是, 当责任者—投稿者检索在高级检索中执行时, 投稿者项目的内容就会与责任者项目者合二为一。

SRS 使用非 MARC 的数据库结构, 并有一个独立的含有全部责任者信息的责任者项目。只有一个责任者的资源按先姓后名的顺序输入; 有两个责任者的资源按第一责任者的姓、名, 然后是第二责任者的名、姓的顺序输入; 有 3 个或多于 3 个责任者的资源如此类推.....。

3.2.8 投稿者

投稿者的信息通常是模糊不清的或是不完整的。遇到这样的情况, 就会输入“没有有关信息”的语句。有些强调某个人或机构在被编目资源生产过程中的作用的说明中就含有投稿者信息。较理想的款目格式来自资源中说明功能或涉及范围时附带的有关个人姓名或机构名称。典型的用来描述投稿者项目的语句有: “由...编辑”、“由...插图”、“由...赞助”、“由...主持”以及“由...维护”。

3.2.9 出版者

SRS 编目人员没有按照 AACR2R 1.4D2 或 1.4D4 中关于编目人员必须使用最简短的格式来描述出版者的规定。他们使用最完整的出版者名称, 加上括号里通用的首字母缩略语来进行著录。例如: “GPO”不会被标记“U S GPO”, 而是“U. S. Government Printing Office (U S GPO)”。在 SRS 完全描述屏幕中, 全称和缩语都会显示出来。如果这个项目中包含全称和缩语, 高级检索中的出版者检索就通过出版者名称或它的缩语来查找记录。

考虑到也许很多用户都不一定知道“s n”的含义, SRS 也没有遵循 AACR2R 1.4D7 中关于出版者不详时使用“s n”的规定。在这种场合下, SRS 使用

“没有有关信息”的语句。

在 AACR2R 1.4D9 中定义了何为未出资料,并规定编目人员在出版者和日期字段中不输入有关未出版资料的任何信息。南希·奥尔森(Nancy Olson)建议:因特网上所有的被挑选来进行编目的资源都应被看作是已经公开出版了的。这样就避免了如何界定什么是已经出版了的因特网资源的问题。SRS 在编目时,把所有在 ISP 出版物中被评述过的资源,以及随后的 SRS 中的资源都看作是已经公开出版了的资源。

3.2.10 资源定位

资源定位的项目是根据美国信息科学研究所(Information Science Institute)的 RFC 1591“域名系统结构与授权”(Domain Name System Structure and Delegation)中关于目前 6 个最高域名的规定:(1).com: commercial, (2).edu: educational, (3).gov: governmental, (4).mil: military, (5).net: network, (6).org: organizational。不在上列域名的资源,如“.k12”、“.ac”和“.co”等,根据资源出版机构及其域名,已经被分配给相应的资源定位。例如:美国信息科学研究所的 RFC 1480“美国域名(The US Domain)”使用“.k12”来标识美国从幼儿园到 12 年级的基础教育。SRS 把这类资料被归到教育类。此外,“.ac.uk”和“.co.uk”是指英国的高等院校(.edu)和商业(.com)网站,并已归到相应的类别。必须指出的是,一个资源只分配给惟一一个资源定位。资源定位只是每一个具体的机构或组织所属类型的一个说明,并不一定关乎资源内容本身。高级检索的用户能够把检索限制在一个定位。这个选项用在附带另外一个检索条件的逻辑和最为有效,有时也可以单独使用。

3.2.11 语言

如果被编目资源中有某一种语言的实际信息内容的话,编目记录就会有所反映。每一种可用的语言都用英语显示。从任何一个 SRS 完全描述屏幕,用户都能够直接连通所有已经被著录了的含有某一种语言的资源。用户如果想寻找某一种具体语言的资源的话,也可以在高级检索里设定一些限制来检索。

3.2.12 主题词

为了给用户提供尽可能多的入口检索到所需资源,这个项目至多能够分配给每个资源 5 个主题词。用 LCSH 主题词浏览是 SRS 4 种主要访问方式之

一,用户可以从按字母的顺序排列的主题词清单来浏览已被完全编目的资源。这种访问方式能显示每一个资源的主题词,以及资源题名及其 URL。这里显示的题名中含有一个直接连通这个网站的超文本链接,和一个连通该资源的 SRS 完全描述屏幕的超文本链接。

实际上,LCSH 存在一些仍未解决的问题。传统的图书馆也许会利用 655 字段来标识一本著作的格式—类别(FORM /GENRE),也会使用主题词“因特网(计算机网络)”来著录一本有关因特网的著作(INTERNET-BASED WORK)。由于 SRS 专门收集有关因特网方面的资源,因此主题词“因特网(计算机网络)”不适用于指明格式(FORM)。SRS 使用 Digital video(数字视频)、short films(短片)、Plot-your-own-stories(标绘个人故事)等来注明格式。然而,如果资源本身就是关于因特网的话,SRS 也会使用“因特网(计算机网络)”来作主题词。

3.2.13 LC 分类号

根据 LC 分类法,SRS 把所有已经编目的资源大致分为 21 个学科类目,每一个学科都分配了 1~3 个字母分类号来表示 LC 分类法中不同主题范畴,组成分层树或分类系统。从每一个学科类目都能够链接到该主题范畴所在子部的目录。例如:用户从 SRS 主页选择社会科学(LCC 分类号 H-HX),就会链接到社会科学子部的所有 16 个主题的目录,从“HA 统计学”到“HX 社会主义和共产主义”。此外,每一个子部都有一个超文本链接连通已经被著录了分类号的资源目录。如果分类号没有超文本链接,则表明这个分类号还没有被分配给所在子部的已经编目的资源。就是说,当一个分类号被分配给一个新编目的资源之后,相应的分类号链接就会被激活。已经被著录了分类号的资源以一个简要描述格式显示,这个格式包括题名、主要 URL、侦察报告摘要(至多 40 字),以及一个连通该资源完全描述屏幕的链接。

SRS 对 LCC 分类表做 3 个方面的改动:(1) SRS 中被编目的资源全部都只有 LCC 分类表的字母分类号,没有分类号之后的扩展数字。(2) 为了更容易访问,许多资源都被著录了两个不同的分类号。例如,老虎映象服务(the TIGER Mapping Service)网站(<http://tiger.census.gov>)同时包含有统计数据 and 地图,因此这个资源就被著录了两个分类号:统计学(HA)与制图学(GA)。这样就能够便于用户识别和进行交叉学科的一系列检索。(3) 有选择性地给

一些LC的注释增加一些说明。例如:注释“耳、鼻、喉”,为了让一些对科学术语不熟悉的用户也能够检索,SR S就增加说明为“耳鼻喉科学”(分类号RF)。

在每一个大学科显示屏幕的底部,用户都可以选择浏览该学科内所有资源题名的完整目录。例如:用户能够看到“所有社会科学题名”的链接,即意味着用户可以浏览分类号以“H”开头的所有资源的目录。

4 有待解决的问题

4.1 缺乏统一标准的因特网编目规则

目前,编目工作人员主要是根据一些非官方的准则来进行著录,如南希·奥尔森的“因特网资源编目”会议演讲、专题讨论以及一些在电子邮件讨论小组的共享输入数据,如AUTOCAT、NTERCAT和Meta2,以及一些印刷版出版物。

随着更多的机构把因特网资源纳入到在他们的网上目录里,有关主题词和细目(格式)子部的需要日益增加。美国国会图书馆已经尝试提出用主题词(如“信息检索系统”、“电子期刊”)描述因特网资源的需要。尽管如此,随着电子信息资源的增长,图书馆提供通往新技术和资源(如“网站浏览器”、“元站点”的主题格式)检索的需求也会增长。

同样需要对ZA子部LCC分类法中的信息资源的分类表做进一步的描述和确定,从而确保因特网资源编目有关规则的一致性。

4.2 难以确定日期信息

日期项目给因特网资源的编目人员提出了另外一个挑战。因为存在不同的标准,如AACR2R、“因特网资源编目”和都柏林核心等,但似乎没有一个能够提出令人完全满意的处理模糊日期信息的解决办法。例如:一个资源往往会有多个日期,如资源被最后一次审校的日期、版权日、一份在不同的日期被变为电子版的文件原来印刷版时最初发表的日期等。

参考文献

- 1 Olson, Nancy B., ed. "Cataloging Internet Resources: A Manual and Practical Guide" Dublin, Ohio: OCLC, 1995. <http://www.oclc.org/oclc/man/9256cat/toc.htm>
- 2 Postel, Jon Network Working Group. "Domain Name System Structure and Delegation" RFC 1591. USC In-

- formation Sciences Institute <http://ds.internic.net/rfc/rfc1591.txt>
- 3 Cooper, A. and J. Postel Network Working Group. "The US Domain" RFC 1480. USC/Information Sciences Institute <http://ds.internic.net/rfc/rfc1480.txt>
- 4 Alvestrand, Harald "Tags for the Identification of Languages" IETF RFC 1766. Internet Engineering Task Force <http://ds.internic.net/rfc/rfc/1766.txt>
- 5 National Information Standard Organization (U. S.). Codes for the Representation of Languages for Information Interchange (NISO standard Z39.53) Bethesda, MD: NISO Press, 1994. <http://www.niso.org/pc-codes.htm#z3953>
- 6 Glassel, Aimee D. & Wells, Amy Tracy. Scout Report Signpost: Design and Development for Access to Cataloged Internet Resources. Journal of Internet Cataloging, 1998, 1(3): 15~45
- 7 Calcari, Susan. The Internet Scout Project Library. Hi Tech, 1997, 15(3-4): 11-18
- 8 Johns, Cecily. Cataloging Internet Resources: An Administrative View. Journal of Internet Cataloging, 1997, 1(1): 17-23
- 9 Neumeister, Susan M. Cataloging Internet Resources: A Practitioner's Viewpoint. Journal of Internet Cataloging, 1997, 1(1): 25-45
- 10 <http://www.signpost.org/>
- 11 <http://scout.cs.wisc.edu/scout/>
- 12 <http://purl.oclc.org/DC/about/element-set.htm>

潘燕桃 硕士,中山大学信息管理系讲师,通讯地址:广州市。邮编 510275。

(来稿时间: 1997-07-20。编发者: 徐苇)