

国际标准文献检索平台的比较及启示

刘佳 钟永恒

【摘要】 文章对 ISO、IEC、ITU、CEN 等主要国际标准文献检索平台进行了系统调研, 并从收录内容、检索方式、著录方式及检索结果四个方面作了分析、评价与比较, 以使用户选择使用。最后, 针对我国标准检索平台重质量轻数量、检索字段少、检准率和检全率低、收录不全、更新慢等不足给出了相关建议, 包括提高著录质量、增加检索字段、提供多语言检索和检索结果的多种排序方式、提高时效性和连续性、整合标准文献。

【关键词】 标准文献 检索 ISO IEC ITU CEN

Abstract: This article analyzes, evaluates and compares the retrieval systems of international standard document, ISO, IEC, ITU and CEN, in terms of content, retrieval style, catalogue information and result, which helps users select and use retrieval systems conveniently and efficiently. And then in view of the weakness of domestic retrieval systems of international standard document, such as quality over quantity, few search fields, poor precision and recall, incomplete content, and delayed update, this article discusses the related measurements, including improving catalogue information quality, increasing search fields, providing multi-language search and different sort, raising the timeliness and continuity and integrating all kinds of standard documents.

Key words: standard document retrieval ISO IEC ITU CEN

标准文献是指由技术标准、管理标准、工作标准及其他具有标准性质的类似文件所组成的一种特种文献。它既具有一般科技文献的作用, 又具有法律效力, 是人们从事科研、生产、设计和检验所使用的技术依据, 也可以直接应用于生产、管理、贸易, 是科技信息检索中不可缺少的内容^[1]。标准文献的数字化, 使标准文献检索变得更加简单、快捷。目前, 国内已经建设了中文标准文献检索平台, 如国家标准馆建设的中国标准服务网、国家标准管理委员会建设的标准检索平台等均可以检索国内外的标准文献。但是由于国外标准从翻译到出版所需时间较长, 以及标准著录格式的不统一, 标准信息用户不能找到最新的和真正需要的标准信息和数据。所以, 在实际工作中, 用户需要将中文标准检索平台和原文标准检索平台配合使用, 才能解决上述问题。本文选取国际标准化组织 (ISO)、国际电工委员会 (IEC) 和国际电信联盟 (ITU)、欧洲标准化委员会 (CEN) 等 4 个国际标准化组织的原文标准检索平台进行分析、比较与评价, 方便用户在标准检索时选择使用, 也为我国标准文献检索平台的建设提供参考。并且, 以国外标准组织的检索平台作为标杆, 对国内标准文献检索平台的建设给出了相关建议。本文选择的标准检索平台是标准信息用户最常用的标准文献信息源, 也是国际标准化组织依托自身的资源优势和优势自建的, 在标准信息和数据的揭示上更加全面、真实。

1 国外标准文献的主要检索平台

国外标准文献检索平台发展非常迅速, 各级标准化组织均建立了自己的标准检索平台, 如国际标准化组织 (ISO)、国际电信联盟 (ITU)、国际电工委员会 (IEC) 等国际标准化文献检索平台, 欧洲标准化委员会 (CEN)、泛美技术标准委员会 (COPANT)、太平洋地区标准大会 (PASC) 等区域标准文献检索平台, 美国国家标准学会 (ANSI)、英国标准化学会 (BS)、德国标准化学会 (DIN)、日本工业标准 (JIS) 等国际标准化文献检索平台^[2]。其中, ISO、ITU、IEC 并称三大国际标准化机构, 在国际标准化活动中占主导地位, 是制定国家标准的重要参考, 其标准检索平台是标准信息用户常用的查找最新的信息和数据的重要来源。本文选择 ISO、IEC、ITU、CEN 等 4 个主要标准文献检索平台, 作为国际或区域标准文献检索平台的代表。这 4 个标准检索文献平台并非相互对立, 可以交叉使用。

ISO 是 1947 年成立的国际标准组织, 宗旨是消除不同国家、地区的相同技术非协调标准形成的技术贸易障碍。ISO 至今已经制定 18 500 项标准, 每年公布约 1 100 项新标准, 标准内容涉及电器、电子领域以外的所有技术领域。技术委员会 (TC)、分技术委员会 (SC) 和工作组 (WG) 负责具体的标准制定工作。有两种途径获取 ISO 标准文献信息, 一是标准目录 (ISO Catalogue) 方式浏览, 按照技术分类代码 (Browse ISO Standards by ICS) 浏览各技术领域的标准制定情况, 或者按照 ISO 标准委员会 (Browse ISO Standards by TC) 浏览各技术委员会的标准制定情况; 二是通过 ISO 标准文献检索平台的搜索引擎快速查询标准文献^[3]。

1.2 IEC

IEC 成立于 1906 年, 是世界上成立最早的国际标准化组织, 负责编制出版电气、电子及相关技术的国际标准。IEC 的标准制定工作由 TC、SC、WG 负责, 其技术专家来自于产业界、政府、测试和研究实验室、学术界和消费者。1996 年, IEC 与 CENELEC 在德国签署了德瑞斯顿合作协议 (Dresden Agreement), 由 CENELEC 制定统一的 IEC 范围外的欧洲电工标准。IEC 的 Web 上提供新闻专栏、报道最新的消息, 包括会议预报、2 周内最新出版的标准^[4]。在 IEC 的 Web 上有两个检索界面, 一种是从主页进入的 Advance Search 检索界面, 可用关键词、标准编号、委员会名称及代号、出版日期、标准状态等进行检索; 一种是从 Websotre 进入的 Search & buy IEC standards 检索界面, 利用标准编号、委员会名称及代号、ICS 号、出版日期等进行检索^[5]。

1.3 ITU

1865 年, 欧洲的 20 个成员国在巴黎签署了世界上第一部《国际电报条约》 (International Telegraph Convention), 并成立了国际电报联盟 ITU (International Telegraph Union)。1932 年, ITU 更名为国际电信联盟, 即 International Telegraph Union (ITU)。ITU 标准是信息通信技术网络的基础, ITU 的研究领域分为: 无线通讯研究领域 (ITU-R)、电信研究领域 (ITU-T)、电信发展研究领域 (ITU-D)。ITU 的 Web 可以按类浏览 (Publications by Sector) 出版物, 或者利用搜索引擎检索整个网站中的网页信息^[6]。

1.4 CEN

CEN 成立于 1961 年, 宗旨是促进成员国之间的标准化协作, 制定本地区需要的欧洲标准 (EN, 除电工行业以外) 和协调文件 (HD)。截止到 2010 年, CEN 已经出版了 12 000 件以上的欧洲标准。CEN 的工作内容包括 EN 标准 (ENs)、EN 预标准 (ENVs)、技术规格 (TSs)、技术报告 (TRs)、CEN 报告 (CRs)、CEN 手册 (CGs)、CEN 工作组协议 (CWAs)^[7]。

2 国外标准文献检索平台的比较及特点

2.1 收录内容

各种标准文献检索平台收录的标准文献不同, 主要表现在内容层次、语言种类、技术领域、标准种类等 4 个方面, 如表 1。英语、法语是最常用的语言, 是各类标准文献都采用的语种; 其次是俄语、西班牙语、阿拉伯语; 汉语虽然是世界上使用人数最多的语言, 但是仅 ITU 标准检索平台提供汉语类标准文献。ISO、IEC、ITU、CEN 这 4 个标准检索平台覆盖的技术领域各异, 在标准信息检索时可以选择使用。ITU 覆盖的标准种类主要是 TRU-T 与 ITU-R 等有效标准和被替代标准。

表 1 各平台收录内容的比较

平台	ISO	IEC	ITU	CEN
内容层次	国际	国际	国际	区域
语言种类	英语、法语、俄语、西班牙语、阿拉伯语	英语、法语、俄语、西班牙语	英语、法语、俄语、西班牙语、阿拉伯语、汉语	英语、法语、德语
技术领域	电子类除外	电工与电子技术	信息通信	电工领域除外
标准种类	有效标准、废除标准、推荐标准、技术报告、标准草案	有效标准、批准标准、推荐标准、标准草案	有效标准、被替代标准	有效标准、技术规格、技术报告、工作组协议

2.2 检索方式

2.2.1 检索字段

标准编号、标准名称、时间范围、标准制定委员会代码、标准种类是基本的检索字段, 除 CEN 外, 各平台均有这 5 个字段。同时, 平台检索字段的具体设置有所差异, 如表 2。ISO 平台的检索字段设置最全面, 其中 ISO 标准状态代码检索字段是 ISO 特有, 共分为 9 个, 分别是: 00—初始阶段, 10—提议阶段, 20—转变阶段, 30—委

员会讨论阶段, 40—公示阶段, 50—批准阶段, 60—出版阶段, 90—复审阶段, 95—撤销阶段, 9 个阶段还包括细分的子阶段^[8]。ISO 和 IEC 均提供标准全文检索和最新标准检索。ISO 最新标准查询分为昨天、本周、上周、本月、上月、本年度、上年度, 较其他标准检索平台更详细。IEC 最新标准查询分为昨天、上周、上两周、上月、上年。CEN 的检索字段除表格中所述的字段外, 还可以选择英、法、德等 3 种语言分类检索。

表 2 各平台可检索字段比较

可检字段 标准名称	标准 编号	标准 名称	摘要	主题词 关键词	被替代 标准	ICS 号	时间 范围	委员会 代号	最新 标准	标准 全文	标准 种类	标准 状态
ISO	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓
IEC	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	×
ITU	✓	✓	×	×	×	×	✓	✓	×	×	✓	✓
CEN	✓	✓	×	✓	×	✓	×	✓	×	×	✓	✓

2.2.2 检索功能

ISO、IEC、ITU、CEN 检索均支持截词检索 (*)、精确检索 (双引号) 和布尔逻辑检索。逻辑检索词包括与、或、非。

2.3 著录方式

标准文献基本的著录信息包括标准编号、标准名称、出版日期、语种、委员会代码和标准状态等。标准编号是标准文献的唯一标识符, 其格式是“标准代号+顺序号+年度号”。ICS 号是指国际标准分类号, 通常 ICS 号分为 40 个大类 (一级类目), 其组成为三层, 第一级由两个数字组成, 第二级由三个数字组成, 第三级由两个数字组成, 各级之间以“.” 隔开。除基本信息以外, ISO 标准著录提供关联标准信息, 包括修改件、技术勘误、修正标准及其他相关标准, CEN 标准著录提供与欧盟公报 (OJEU) 的引用关系, 这些信息体现了知识的流动方向, 是了解技术发展过程的重要信息。CEN 著录的特点主要体现在语种和日期上, CEN 同时提供英、法、德 3 种语言的著录信息, 日期著录不仅包括出版日期 (DOP)、有效日期 (DAV), 还包括公开日期 (DOA)、批准日期 (DOR)、撤销日期 (DOM)。此外, IEC 标准提供标准正文预览, ITU 提供标准文献目录信息。见表 3。

表 3 各平台标准文献著录格式比较

著录信息 标准名称	标准 编号	标准 名称	出版 日期	版次	文献 格式	价格	页码	语种	摘要	委员会 代号	ICS 号	有效期	标准 状态
ISO	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
IEC	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓
ITU	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	×	✓	✓
CEN	✓	✓	✓	×	×	×	×	✓	×	✓	✓	✓	✓

2.4 检索结果

2.4.1 检索结果的排列方式

各类标准检索平台的检索结果均提供了多种排序方式, 更加方便用户浏览和筛选相关标准文献。ISO 和 IEC 均提供了 6 种不同的排序方式, 但是 ISO 提供的排序方式更加符合标准信息用户的行为习惯。ITU 未提供检索结果的排序选项, 默认按照相关度排序。见表 4。

表 4 各平台检索结果可选的排序方式比较

平台	排 序 方 式
ISO	标准编号、ICS 号、委员会和分委员会、标准/项目所达到的阶段、公布日期、与检索词的相关性排序。
IEC	标准编号、委员会、标准类型、语言、是否提供预览、是否是 CENELEC 标准。
ITU	无。
CEN	标准编号、委员会、项目编号。

2.4.2 检索结果数量比较

不同的标准文献检索平台, 利用相同的关键词检索, 原则上得到的标准文献数量一致。但通过比较后发现, 原文检索平台得到的 ISO 标准文献数量远少于中国标准信息网 (CSSN) 和中国知网国外标准数据库 (CNKI), 而得到的 IEC、ITU、CEN 标准文献数量多于 CSSN 和 CNKI, 具体数据见表 5。

ISO 中不提供关键词检索, 中文检索平台中提供的关键词检索字段是属于著录人员添加的字段, 扩大了检全率和检准率, 所以 ISO 标准在中文检索平台中的检索结果数量远远多于原文检索平台中的检索结果。以现行国际标准 ISO 12871-2010 为例, 该标准的 ICS 号为 67.200.10 (食品大类中食用油和脂肪、含油种子类), 所属技术委员会为 ISO/TC 34 (食品类), 属于食品类标准。在中文检索平台中能通过关键词 food 检索到, 而在 ISO 检索平台不能通过摘要或者题名检索到, 只有利用 ISO 全文检索才能检索到, 而利用全文检索检准率不高。利用 ISO 检索平台检索时, 需要利用 ISO 的题名、摘要、全文检索、技术委员会、ICS 号等综合检索, 保证 ISO 标准的检全率和检准率。

原文检索平台本身收录的标准文献信息比中文检索平台更具有时效性和连续性。中文检索平台收录外文国际标准文献时, 翻译上存在时滞, 并且对同一标准号的多种版本的标准文献收录不全面。

此外, CNKI 和 CSSN 中对于同一检索词、同一类标准文献的检索结果数量存在差异, 原因有两个: 一是 CSSN 中标准更新速度比 CNKI 慢, 以欧洲标准 EN 4604-010-2009 为例, 该标准发布日期为 2010 年 8 月, 利用关键词 light 可以在 CNKI 中检索到, 而在 CSSN 中不能检索到; 二是 CSSN 中收录标准较 CNKI 更全面, 以 EN 2155-5-1989 标准为例, 在 CNKI 只能够检索到 1 件, 而在 CSSN 中能够检索到 11 件, 与 CEN 正式出版的标准 11 件数量相等。

表 5 各平台检索结果数量比较

检索词及标准名称		标准种类	正式公布	撤销/替换	标准草案
食品	ISO		250	82	13
	CSSN (ISO)		1231		
	CNKI (ISO)		1371		
计算机	IEC		221	38	12
	CSSN (IEC)		204		
	CNKI (IEC)		194		
卫星	ITU		32	14	\
	CSSN (ITU)		45		
照明	CEN		455	\	20
	CSSN (CEN)		213		
	CNKI (EN)		240		

注: ①标准检索日期为 2011 年 7 月 22 日; ②中文标准检索平台不提供标准类型分类检索, 检索结果只有标准总数的统计, 包括了现行标准、撤销或替换标准以及标准草案; ③ CNKI 中不提供 ITU 标准文献检索; ④ CNKI 中只提供 CEN 标准中 EN 标准检索; ⑤ CEN 中不提供撤销/替换标准文献检索。

3 对国内标准文献检索平台的启示

3.1 国内标准文献检索平台的建设现状

学者蔡莉静^[9]、王晋生^[10]、李伟华^[11]等对国内标准文献检索平台的建设现状及特点进行了总结。本文通过与国际标准文献检索平台的比较, 发现国内标准文献检索平台主要存在以下不足: (1) 注重收录总量, 忽视著录质量。国内检索平台均收录 ISO、IEC、ITU 和 CEN 等国际或区域标准, 部分检索平台收录了美国国家标准 (ANSI)、美国材料与试验学会标准 (ASTM)、联邦德国国家标准 (DIN)、英国国家标准 (BS)、法国国家标准 (NF) 等国家标准、行业标准。但是对于标准文献著录不全面, 如 ICS 号、关联标准等信息缺失, 部分标准文献未提供中文

标题,且 ICS 号的著录格式不统一。(2) 检索平台的检索字段偏少。国内标准文献检索平台均提供的检索字段有中英文标题、中英文关键词、ICS 号、国家标准分类号、标准编号等,但是都没有标准制定委员会、标准状态、标准种类、最新标准等检索字段。(3) 检准率和检全率低。国内检索平台的标准号字段对于部分外文标准文献不适用,且多数平台不支持布尔逻辑检索。(4) 标准文献收录不全面,对于有多个版次的标准文献只收录了部分。(5) 标准文献更新慢。

3.2 建议

针对以上不足,国内标准文献检索平台可以从以下几方面改进:(1) 注重对原文标准数据的深加工以及著录质量的审查,包括外文标题与摘要的翻译、不同版次标准文献的关联、不同类型标准文献的关联、中文标准分类号与 ICS 号的映射等。(2) 提供检索结果的不同排序方式,如按相关度排序、按标准编号排序、按 ICS 号排序、按标准类型排序、按标准发布日期排序等,方便标准信息用户浏览和筛选真正需要的标准文献。(3) 增加检索字段,如最新标准、标准状态、标准制定的技术委员会、标准的 ICS 号等字段。(4) 提供多种语言检索。中文检索平台一般提供中、英文两种语言检索,但是,未提供其他使用也较多的法语、西班牙语、阿拉伯语等语言的检索方式。(5) 整合各个标准组织的标准文献,特别是重要技术领域的国际标准文献、区域标准文献、国家标准文献。(6) 注重标准文献收录的时效性和连续性。

注释

[1][2][10] 王晋生. 我国技术标准免费检索系统的比较分析. 情报科学, 2006 (5): 736-740

[3] ISO <http://www.iso.org/iso/home.htm>, 2011-07-23

[4] 陈陶, 夏立娟. ISO、IEC、ITU 标准文献的网上检索. 图书馆学研究, 2004 (8): 75-77

[5] IEC <http://www.iec.ch/>, 2011-07-23

[6] ITU <http://www.itu.int>, 2011-07-23

[7] CEN <http://www.cen.eu>, 2011-07-22

[8] ISO http://www.iso.org/iso/standards_development/processes_and_procedures/stages_description/stages_table.htm#s60, 2011-07-24

[9] 蔡莉静, 姚新茹, 闫辉等. 如何有效地利用互联网上的标准数据库. 图书情报工作, 2004 (9): 86-89

[11] 李伟华, 王通, 顾英. 网络标准文献信息资源的分布及检索. 情报探索, 2010 (12): 74-77

刘佳 中国科学院文献情报中心, 中国科学院国家科学图书馆武汉分馆, 硕士研究生。

钟永恒 中国科学院武汉文献情报中心, 中国科学院国家科学图书馆武汉分馆, 研究员。

(上接第 49 页)

5.4 实行分级存储, 确保数据安全

存储系统要设置在线、离线、灾备等多级存储结构。在线存储可满足数据加工、处理及补充上传等工作需求; 离线存储可满足数据的归档/备份等特殊需求; 如条件允许最好设置异地灾备存储功能、满足意外情况下的可用性安全需求。完善多级存储, 并实现所有数据的定期归档、统一迁移、适时备份和还原, 确保数字资源万无一失、安全运行。

注释

[1] 我国数字图书馆标准规范之资源集合描述元数据规范. <http://cdls.nstl.gov.cn/2003/SetMetadata/>: 2011-04-25

[2] 专门数字对象描述元数据规范设计指南. http://cdls.nstl.gov.cn/m/blog/2nd_archive/doc/, 2011-04-25

[3] ZADL 特色数据库评价指标体系, 浙江高校数字图书馆(ZADL)资源发展组

胡绍军 浙江理工大学图书馆馆长助理兼资源建设部主任, 副研究馆员。