

军用信息检索语言易用性研究

赵建国, 周健

(中国人民解放军军事科学院, 北京 100091)

摘要: 概要介绍了军用信息检索语言的发展情况, 论述了体现军用信息检索语言易用性的三个层面, 分析了影响军用信息检索语言易用性的因素, 提出了军用信息检索语言研究方向和军用信息检索语言编管工作发展战略。

关键词: 军用信息检索语言; 易用性

中图分类号: G354

DOI: 10.3772/j.issn.1673—2286.2014.10.003

军用信息检索语言是军事信息资源组织、存储与检索时用来对文献信息内容和信息需求进行主题标引、特征描述和逻辑分类, 进而表达文献信息内容概念及其相互关系的一种规范化的人工语言, 其职能是为信息检索系统提供语言保障^[1]。

1 军用信息检索语言的发展概述^[2-5]

中国人民解放军的信息检索语言编制管理工作以1987年7月“中国人民解放军军用主题词表编制委员会”成立为标志, 已经走过27年的历程。1991年5月, 该委员会改称“中国人民解放军军用主题词表编制管理委员会”, 由单一的“编制”职能扩展到“编制、管理”职能; 根据军队信息化发展需要, 又于2011年5月改称“中国人民解放军军用信息检索语言编制管理委员会”, 工作领域由单一的《军用主题词表》扩展到整个军用信息检索语言。委员会主要负责全军军用信息检索语言的编制、管理、应用和科学研究工作。其办公室设在军事科学院, 具体负责全军军用信息检索语言编制、管理和应用中的计划与组织协调、业务指导、理论研究等日常事务。27年来, 在中央军委和总部首长的关怀下, 在全军军用信息检索语言编制管理委员会和军事科学院的领导下, 在军内外专家和全军各有关单位的大力支持下, 军用信息检索语言编制管理取得了可喜的成绩, 共取得60多项成果, 获得国家和军队科研奖励30余项。

目前, 军用信息检索语言主要包括军用主题检索语言和军用分类检索语言两种类型。军用主题检索语言主要包括1部大型综合性主题词表《军用主题词表》和《军用公文主题词表》、《海军主题词表》、《国防科学技术叙词表》等23部与《军用主题词表》相互兼容、结合使用, 覆盖各军兵种的专业性主题词表, 开创了主题词表应用系列化、兼容化的先河; 各主题词表一般由字顺主表和若干索引附表组成, 其中索引表数量少则一两部, 多则四五部。索引表的多寡一般根据字顺表的具体展示结构、使用对象对词表的可能性需求、编表的时间周期长短等因素, 选择编制其中一部分, 包括范畴表、释义表、词族表、轮排表、汉英表、英汉表等。军用分类检索语言主要包括1部大型综合性分类法《军事信息资源分类法》和80余部专业性分类法、分类表、代码表或分类代码表。各分类表一般由基本大类、简表、详表和若干复分表组成。复分表一般包括综合复分表、军兵种复分表、世界地区复分表、中国地区复分表、国际时代复分表、中国时代复分表、资源类型复分表等。

此外, 还研制开发了智能型的“军用主题词表应用管理系统”和“军用主题词表辅助标引系统”, 在国内率先实现了主题词表应用管理的电子化、智能化; 编修完成18部信息资源管理领域的国家军用标准。这些成果对服务部队信息资源建设、提高信息资源管理利用效能发挥了重要作用, 得到了国内信息管理和标准化领域同行的关注和好评。军用信息检索语言事业从无到有, 从小到大, 有了长足的进步, 技术水平达到国内领先, 为有效

服务部队、机关、院校和科研单位发挥了重要作用。

2 军用信息检索语言的易用性

信息检索语言是一种人工语言,具有词汇管理成本高、标引负担重、结构复杂、不易被标引人员和检索人员掌握等与生俱来的弱点。其易用性是信息检索语言可用性评价最重要的一级指标^[6],是指信息检索语言容易被用户接受和使用的程度。信息检索语言的易用性体现在“快”、“便”、“省”上,即标引和检索速度要快、要方便、成本要低、效益要高^[7]。在降低成本、克服障碍、方便使用方面发挥作用,有利于提高“全”和“准”。提高信息检索语言的易用性就是采用一系列必要的措施和手段,研究用户心理,注重用户体验,降低其使用成本和难度,使其易于被用户接受和使用,从而提高工效,减少错误,提高标引和检索效率。这些措施和手段包括保证检索语言概念体系的丰富性和完整性、重视运用方法性注释、充分显示概念之间的关系、采取分面控制措施、提供多种查词查号手段等等^[8]。具体说来,在词汇或概念的选取标准上,既要保证语法规则的严密性、严格进行标识的规范化处理、保证概念标识含义的明确性,又要选用绝大多数用户可接受的词语或符号做标识,并适当引进自然语言,尽可能多地收录入口词;在词表宏观结构和微观结构的设计上,既要尊重科学、符合逻辑、充分展示,又要详略得当、避免繁琐、易于使用;在词表内容的计算机显示、查询、浏览上,既要尊重用户的使用习惯,又要使词表排列显示清晰明了、易于理解,便于浏览查询;在词表的动态维护更新上,既要随时收集用户的意见与建议,又要及时更新并发布最新的数据。

传统的信息检索语言以印刷版形式,分多卷出版专用工具书。用户需要依靠手工一遍遍反复查阅,查词、查义、查类目,分析文献主题,提炼主题概念,转换主题词、分类号,最后标引主题词、分类号,费时费力,文献标引工作的效率相对较低。我们认为,传统信息检索语言要以用户为中心,进行易用性改造。军用信息检索语言要尽可能人性化,注重用户体验,以满足广大用户的信息需求为宗旨,实现从面向专业人员到面向普通用户的转变。鉴于此,在军用信息检索语言编管实践中,我们把“以人为本”作为编管工作的重要原则,着重加强信息检索语言在标引和检索文献方面的功能,以增强其易用性。

军用信息检索语言的易用性大致体现在以下3个层面。

2.1 对于军用信息检索语言传统印刷版形式的计算机再现

(1) 检索语言在计算机上以类似印刷版的形式再现。逐页浏览,逐词查询,可以更加方便地逐页定位、逐词定位,还可以任意放大或缩小显示字号。可以设计成多种电子图书格式,如网页格式、PDF格式。

(2) 主题词表的字顺主表及各索引附表的多屏幕多窗口再现。在计算机单一屏幕上的多个窗口中,以各自表的印刷版形式显示。可以从任意窗口查询主题词,在多个窗口的各表同步、联动显示,可成倍地提高词表查询效率,极大地提高词表易用性。

2.2 对于军用信息检索语言多维语义信息的深层次揭示

(1) 军用主题词表语义信息的多维度、立体化展示。多屏幕多窗口查询、定位,同步、联动显示、浏览,打破了传统词表的线性流水式显示局限,实现了语义词表从多个维度,全方位立体化展示词表丰富的概念语义信息。

(2) 军用主题词表动态化组版排版,集约化浏览显示。利用计算机技术,屏蔽无关概念语义信息,关注用户主题语义焦点,动态组版排版,形成一个临时性、小型独立词表,实现主题词表集约化显示浏览。排除无关干扰,集中用户精力,提高了词表易用性。

(3) 《军事信息资源分类法》的易用性设计主要体现在多维揭示^[9]上,同时,在类目设置、类名的选择、分类表的显示以及分类表的使用上,为用户提供便利条件,不出现空链、死链^[10]。

a.按照事物的不同属性设置类目,提供从多个角度揭示和检索信息资源的方法。着重体现构成军事科学完整体系的军事理论和军事技术两大组成部分,并将军事理论类目列于前,军事技术类目列于后。对于内容庞杂、类无专属的信息资源,概括为“综合信息资源”,列于最后。类名尽量使用《军语》中的规范用语以及《军用主题词表》中的正式主题词。

b.按照类目之间的内在联系组织类目体系,充分显示一个学科或专业领域中知识门类及其关系的全貌,便于从某学科或专业出发浏览信息资源;完善各种注释,通过缩格、注释等方式准确、充分地显示类目之间的从属、并列、交叉、交替、相关等语义关系,并编制简明易用的

《使用手册》。

c.以网状结构多方位、多层次地揭示知识的内在相关性,并根据用户浏览的需要,灵活地处理知识间的交叉关系。当一个知识门类具有多种从属关系时,采用交叉列类方法,即一个子类可以同时属于多个母类,以充分揭示主题概念之间的联系,既保持学科体系的完整性,又不分散同一内容的信息资源。

d.类目的划分原则上遵循同一个标准,为了多视角地揭示知识的内在联系,类目划分不局限于单一标准的逻辑划分,而是根据需要同时使用若干标准进行划分,建立若干个从属母类的平行子系统。

2.3 对于军用信息检索语言应用不同程度的自动化辅助标引

长期以来,我们试图把大量的词汇控制、转换等工作交给计算机后台去处理,让用户面对界面友好、易学易用的“傻瓜词表”,着力提高军用信息检索语言应用的便捷性和易用性。

(1)文献标引由人工提炼的主题概念自动转换为主题词。第1版的“军用主题词表应用管理系统”是半自动化的计算机辅助标引,即由标引员人工提炼文献的主题概念,再由计算机分析相对规范的主题概念,将其转换为词表中的正式主题词。

(2)文献标引由重点段落自动推荐出标引词。第2版的“军用主题词表辅助标引系统”,较之“军用主题词表应用管理系统”而言,易用性明显提高。用户可以在电子文献中任意选取重点段落,可以是标题、文摘、文首段、文尾段、各章节悬置段,甚至全文。系统自动推荐标引词,供用户选择。该系统基本上属于智能型傻瓜标引器,用户操作基本上仅限于点击鼠标。

(3)受控语言自然语言化,即与自然语言广泛接口,提高词表入口率。为提高转换效率、推荐准确率,需要大幅度提高词表入口率,词表的入口词越多越好,国外词表入口率可高达100%。《军用主题词表》第1版,即1990年版,词表入口率10%(非正式主题词数量/正式主题词数量*100%);修订后的第2版,即2002年版,其入口率增加至50%。计算机信息处理技术以及词表受控语言自然语言化技术的应用使《军用主题词表》和“军用主题词表辅助标引系统”的易用性、实用性明显提高。

3 影响军用信息检索语言易用性的因素

在发展过程中,影响军用信息检索语言易用性效能发挥的薄弱环节主要表现在编管工作中“编”与“管”、“研”与“用”的不均衡等方面。

3.1 “编”与“管”发展不均衡

军用信息检索语言在编制方面硕果累累,但动态管理没有及时跟进。由于军用信息检索语言编管工作网络建设滞后,还没有建立一套快捷、高效的信息发布机制,用户单位对于军用信息检索语言编管工作的需求和反馈不能及时反馈,其实际应用效果参差不齐,难以实施有效跟踪和管控。

3.2 “研”与“用”发展不均衡

军用信息检索语言的推广应用有些方面还不及、不到位,实际效果还没有准确评估。如何改变信息检索语言枯燥乏味的形象,使机关、院校和部队用户真正体验到各种易用性检索语言工具的乐趣,进而使军用信息检索语言编管工作有效贴近强军目标、服务军事斗争准备,还需要进一步探索。此外,现有各种军用信息检索语言编制管理系统、辅助标引系统在技术上国内领先,这些易用性系统如何更多更快地走出军队,在国家信息化建设中一显身手,实现军地共享,也是值得认真考虑的事情。

此外,在易用性方面,《军用主题词表》中的附表、《军事信息资源分类法》中的复分表以及国家军用标准《军用主题词组织机构名称缩略规则》的计算机应用,如人名、地名、著作名、组织机构名、武器型号名等,都是亟待解决的问题。

4 军用信息检索语言研究的方向

目前,全军军用信息检索语言编管会办公室正在研制开发的项目有4项。

4.1 文献智能化辅助分类研究

我国文献辅助分类、自动分类研究已经多年,一般为聚类研究,或者是几十个、至多是一两百个类目的分类标引。而《军事信息资源分类法》这样八千多个类目的计算机自动分类实用系统还很少见诸报端。

4.2 《军用主题表》与《军事信息资源分类法》相互融合

编制《军用分类主题映射表》，建立类目对应主题词、类目标注主题词、扩展类目主题词、类目隶属主题词等映射关系；为分类号与主题词一体化同步标引和检索服务，也为计算机上的一体化同步应用做数据准备。

4.3 通过建立百万条军事词汇大词库提高词表入口率

建设百万条军事词汇大词库，使综合性主题词表与专业性主题词表更加紧密地结合与应用，并适时修订《军用主题词表》，使词表入口率提高到100%，提高主题词转换效率、转换准确率，使主题词表更加易于使用。

4.4 建立检索语言分类号与主题词一体化同步标引和检索系统

文献标引过程中，分类号与主题词相互借鉴、相互印证，进一步提高主题词转换效率、转换准确率，使主题词表更加实用。

5 军用信息检索语言编管工作发展战略思考

经过27年的发展，我们感到，军用信息检索语言编管工作必须开阔眼界，与时俱进，真抓实干，牢牢把握军队信息化建设大方向，与部队信息化建设需求接轨。我们对下一步工作的总体考虑是：聚焦强军目标，立足我军信息检索语言研究现状，瞄准国内外信息检索语言发展前沿，顺应信息检索语言发展趋势，紧贴军队信息化建设实际，深入开展理论研究，大力推动技术创

新；以军用信息检索语言编管服务平台建设为契机，扎实推进军用信息检索语言网络化、电子化、智能化、易用化，不断挖掘军用信息检索语言成果推广普及应用新亮点；充分发挥军队信息检索语言编制管理常设机构的优势，继续发扬艰苦创业、与时俱进、开拓进取的精神，让“小核心、大外围”、“小机构、大科研”的组织模式和“强强联手”、“军地结合”的工作机制发挥更大的作用；同时，科学规划、统筹管理，正确处理好“编”与“管”、“研”与“用”的关系，努力促进军用信息检索语言在我军标准化、信息化建设以及作战信息系统中发挥更大效益，探索军地科研成果共享，早日实现军用信息检索语言编管工作新的转型。

参考文献

- [1] 张琪玉,侯汉清.情报检索语言实用教程[M].武汉:武汉大学出版社,2004.
- [2] 武天敏.国家军用标准《军事信息资源分类法》通过评审[N].解放军报,2005-04-23(1).
- [3] 包国俊.军用主题词表辅助标引系统在全军推广应用[N].解放军报,2006-07-08(1).
- [4] 包国俊,武天敏.我军信息检索语言编管工作成果丰硕[N].解放军报,2007-08-20(1).
- [5] 武天敏,包国俊,郭瑞宏.在浩瀚信息海洋“导航”:与全军军用主题词表编管会办公室主任赵建华对话录[N].解放军报,2007-08-23(12).
- [6] 赖茂生,王婧,麦晓华.检索语言可用性评价初探[J].情报理论与实践,2012(8):65-69,96.
- [7] 张正强.对情报检索语言易用性问题的思考[J].图书馆,1994(2):29-30.
- [8] 张琪玉.张琪玉文集[M].北京:国家图书馆出版社,2014.
- [9] 赵建华,陈庆平,周健,等.《军事信息资源分类法》的总体设计与技术特色[J].中国索引,2005(4):34-39.
- [10] 刘巧英.传统文献分类法在网络中的优势与缺陷及其发展[J].图书馆学研究,2004(5):51-53.

作者简介

赵建国,男,1960年生,军事科学院全军军用信息检索语言编管会办公室研究员,研究方向:信息组织、信息检索语言。

周健,男,1963年生,军事科学院全军军用信息检索语言编管会办公室副研究员,研究方向:信息组织、信息检索语言, E-mail: zj1963@tom.com。

Study on the Usability of PLA Military Information Retrieval Languages

ZHAO JianGuo, ZHOU Jian

(PLA Academy of Military Science, Beijing 100091, China)

Abstract: This paper briefly introduces the development of PLA military information retrieval languages, discusses the three levels of the usability of military information retrieval languages, analyzes the influencing factors for the usability of the work, and puts forward the research directions of military information retrieval languages and the development strategy of the compilation and management of PLA military information retrieval languages.

Keywords: Military information retrieval languages; Usability

(收稿日期: 2014-07-11)