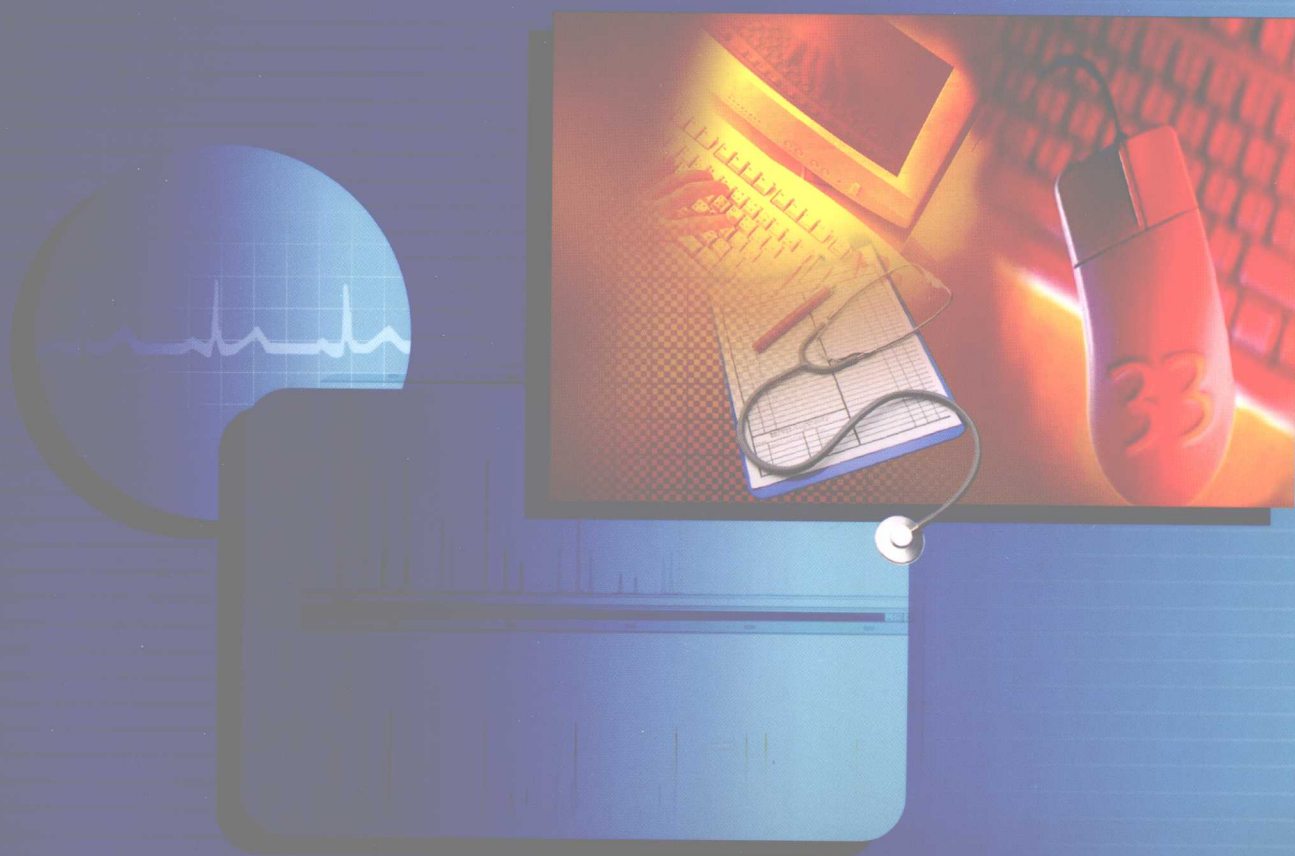


全国高等学校医学规划教材
(供信息管理与信息系统专业用)

医学文献主题标引

主编 肖晓旦 张士靖



高等教育出版社
Higher Education Press

全国高等学校医学规划教材
(供信息管理与信息系统专业用)

医学文献主题标引

主 编 肖晓旦 张士靖

副主编 沈秀丽 侯跃芳 任淑敏



高等教育出版社
Higher Education Press

内容简介

本书为高等学校信息管理与信息系统(医学信息学)专业的规划教材之一,在阐述情报检索语言基本概念、类型及其作用的基础上,通过揭示 MEDLARS 检索语言的构建基础与特点,探讨了特征词、副主题词和各类生物医学文献主题标引的规则和方法,并对中医药文献的标引规则和方法进行了讨论。

本书除了可作为全国高等学校医学信息学本科生的教材外,也可作为医学期刊编辑、生物医学文献检索系统与数据库构建、网络生物医学信息资源组织、图书馆生物医学文献主题目录和主题索引编制人员的重点参考书。

图书在版编目(CIP)数据

医学文献主题标引/肖晓旦,张士靖主编. —北京:
高等教育出版社,2008.10

供信息管理与信息系统专业用

ISBN 978-7-04-023635-4

I. 医… II. ①肖…②张… III. 医学-文献-主题标
引-高等学校-教材 IV. G254.36

中国版本图书馆 CIP 数据核字(2008)第 120069 号

策划编辑	冯娟	责任编辑	王友富	封面设计	张楠	责任绘图	宗小梅
版式设计	马敬茹	责任校对	王超	责任印制	尤静		

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
总 机 010-58581000

经 销 蓝色畅想图书发行有限公司
印 刷 化学工业出版社印刷厂

开 本 850×1168 1/16
印 张 18
字 数 520 000

购书热线 010-58581118
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landrac.com>
<http://www.landrac.com.cn>
畅想教育 <http://www.widedu.com>

版 次 2008 年 10 月第 1 版
印 次 2008 年 10 月第 1 次印刷
定 价 28.60 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 23635-00

编委名单

主 编 肖晓旦 张士靖

副主编 沈秀丽 侯跃芳 任淑敏

编 者 (按姓氏拼音字顺排)

侯跃芳 中国医科大学

李玉玲 吉林大学

任淑敏 济宁医学院

沈秀丽 吉林大学

孙凤梅 滨州医学院

王秀平 山西医科大学

肖晓旦 中南大学

张士靖 华中科技大学

秘 书 刘海霞

前言

医学文献主题标引作为一门最能体现医学信息学专业特色的核心课程,其主要任务是通过本课程的学习,使学生在对情报检索语言尤其是叙词法的基本原理与方法全面了解的基础上,掌握生物医学文献主题标引的规则与技巧,达到标引的准确性、符合性和一致性。

美国国立医学图书馆在生物医学文献信息的组织、处理与交流过程中,利用叙词法的原理与方法研制了著名的《医学主题词表》及其标引规则,其词表和规则为许多国家和地区所采用,成为全世界生物医学文献主题标引与检索的通用词表与规则。从20世纪80年代起,我国医学图书情报界就引进和利用其词表、规则和标引技术来组织和处理生物医学文献资源,积累了较丰富的标引和检索经验。中国医学科学院医学信息研究所从1984年起,连续编译了三版《医学主题词注释字顺表 MeSHAAL(英汉对照表)》,中国第一军医大学编译了《医学主题词树状结构表(英汉对照表)》,针对 MeSHAAL 对中医药文献主题词收录不充分的缺陷,中国医学科学院医学信息研究所编制了《中国中医药学主题词表》并制定了相应的标引规则,形成了我国较完善的生物医学文献标引与检索体系。

本书从情报检索语言尤其是叙词法的基本原理与方法着手,对美国国立医学图书馆编制的《医学主题词注释字顺表》及其标引规则、中国医学科学院医学信息研究所编制的《中国中医药学主题词表》及其标引规则进行了系统的阐释,为生物医学文献资源的主题标引提供了理论和实践指南。

本书既是医学信息学专业本科生学习医学文献主题标引的教科书,又是传统或网络生物医学文献主题组织的工具书,具有较强的实用性。为了加强对学习者动手能力的培养,本书列举了大量的标引实例以诠释标引原理与规则,并在章后附有思考题或标引题。本书编者在阅读和参考大量专家学者的相关文献基础上,力求反映 MEDLARS、MeSHAAL 和中国中医药学文献标引检索系统的最新变化和进展。

针对大多数医学信息学本科专业没有单独设置情报检索语言这门课程,而这方面的理论基础对于正确理解和掌握医学文献主题标引方法与技巧又是十分必要的这一实情,本书系统地介绍了情报检索语言,尤其是叙词检索语言的性质与特点,并对自动标引的原理进行了探讨。

本书在介绍情报检索语言和 MEDLARS 检索语言的体系结构的基础上,探讨了生物医学文献主题标引的一般原则与方法、医学主题词注释字顺表及其标引工具书结构与使用方法,从 MEDLARS 副主题词标引规则着手,对83个副主题词按其族性结构逐一进行了讨论,并对各类生物医学文献主题词的标引方法与规则,尤其是A类解剖学主题词、B类有机体主题词、C类疾病类主题词、C4类肿瘤主题词、D类药品和化学物质主题词、E类技术类主题词、中医类主题词以及其他类主题词的标引方法与规则进行详细的讨论。各院校可根据自己的实际情况,调整讲授内容。

本书在共同讨论的大纲基础上进行编写,各章的编写者分别为:第一、八、九章和第三章(部分),肖晓旦;第二、四、六、十一、十三章,张士靖;第三章(部分)和第十二章,沈秀丽、李玉玲、孙凤梅;第五、七章,侯跃芳;第十章,任淑敏;第十四章,王秀平。全书由肖晓旦、张士靖负责统稿。

本书在编写和出版过程中参考了大量学者的研究成果,得到了许多领导及专家的大力支持,在此表示衷心的感谢。

由于编写水平有限,难免有错漏之处,敬请专家、学者和广大读者批评指正。

编者

二〇〇八年二月

目 录

第一章 绪论	1	第三节 我国医学文献主题标引工作与 MEDLARS 标引语言	71
第一节 主题标引课程的意义与任务	1	第八章 医学主题词表	74
第二节 情报检索语言概述	2	第一节 主题词表的概念、作用和结构	74
第三节 情报检索语言构建基础	4	第二节 医学主题词注释字顺表	75
第二章 情报检索语言的类型	8	第三节 医学主题词树状结构表	88
第一节 情报检索语言的种类	8	第四节 在线 MeSH 表	92
第二节 等级体系分类语言	9	第九章 其他标引工具书	97
第三节 组配分类语言	13	第一节 技术札记——MEDLARS 标引指导	97
第四节 标题词语言	14	第二节 技术札记补篇	98
第五节 单元词语言	16	第三节 医学主题词表——补充概念记录 (补充化学记录)	101
第六节 叙词语言	18	第四节 医学主题词轮排表	103
第三章 自然语言在情报检索中的应用	26	第五节 医学主题词词义注释	104
第一节 自然语言的应用是当今情报检索领域 的一种重要发展趋势	26	第十章 特征词	106
第二节 关键词法概述	27	第一节 特征词概述	106
第三节 关键词法在生物医学文献信息资源 组织中的运用	31	第二节 特征词的标引规则	107
第四节 自动标引	33	第三节 文献类型特征词	112
第四章 标引的基本知识	40	第十一章 副主题词	116
第一节 标引概述	40	第一节 副主题词概述	116
第二节 标引的原理	42	第二节 副主题词字顺表	119
第三节 标引深度和标引等级	45	第三节 副主题词树状结构表	121
第五章 主题标引的原则与步骤	49	第四节 副主题词标引原则	123
第一节 主题标引的一般原则	49	第十二章 各类主题词的标引	126
第二节 主题标引的步骤和方法	50	第一节 A 类 解剖学主题词标引	126
第六章 概念组配和组配标引	60	第二节 B 类 有机体主题词标引	132
第一节 概念组配的原理	60	第三节 C 类 疾病主题词标引	137
第二节 概念组配方法和原则	62	第四节 C4 类 肿瘤主题词标引	145
第三节 组配标引的类型和常见错误	63	第五节 D 类 化学物质和药物主题词 标引	149
第七章 我国医学文献主题标引与 MEDLARS 标引语言	65	第六节 E 类 诊疗技术及设备主题词标引	156
第一节 美国国立医学图书馆与 MEDLARS	65	第七节 F 类 心理学和精神病学主题词 标引	161
第二节 MEDLARS 标引语言的结构体系	67	第八节 G 类 生物科学主题词标引	163
		第九节 H 类 自然科学主题词的标引	167

第十节 I类 社会科学主题词的标引	169	第四节 与化学有关的副主题词	214
第十一节 J类 工艺、工业、农业、食品类 主题词的标引	173	第五节 与药理学有关的副主题词	218
第十二节 K类 人文科学类主题词的标引	175	第六节 与诊断有关的副主题词	224
第十三节 L类 信息科学主题词的标引	177	第七节 与治疗有关的副主题词	226
第十四节 M类 名和职业名称主题词的 标引	180	第八节 与生理学有关的副主题词	233
第十五节 N类 医疗保健主题词的标引	185	第九节 与统计学有关的副主题词	240
第十六节 Z类 地理名称主题词的标引	190	第十节 与组织和管理有关的副主题词	244
第十三章 副主题词各论	194	第十一节 其他副主题词	248
第一节 与解剖学有关的副主题词	194	第十四章 中医药学文献主题标引	262
第二节 与病因学有关的副主题词	200	第一节 中医药学文献主题标引概述	262
第三节 与分析有关的副主题词	211	第二节 中医药学文献主题词标引规则	267
		第三节 中医药学文献副主题词标引规则	272

内容提要

医学文献主题标引是医学信息学本科专业一门重要的专业课程,如何掌握生物医学文献主题标引的规则与技巧,科学、合理地组织生物医学文献信息资源,有效地建立情报检索系统,进行情报检索,是医学信息学教育的重要任务之一。本章主要介绍本课程的意义与任务、情报检索语言基本概念和构建基础。

医学文献主题标引是医学信息学本科专业一门重要的专业课程。它主要解决科学、合理地组织生物医学文献资源的问题,即按照一定的科学原理、规则和方法使大量的、无序的生物医学文献信息资源成为有序的信息资源,它是解决医学文献信息资源需求与利用之间的矛盾,有效地进行医学文献信息资源管理和文献信息资源增值的基础。从情报学角度看,它属于情报加工、存贮的范畴,是开展情报分析研究与情报服务的前提条件。本书以美国国立医学图书馆(NLM)创建的大型生物医学文献集成系统——医学文献分析与检索系统(MEDLARS)的标引方法与技术为基础,全面而系统地介绍生物医学文献主题标引的基本原理、规则与方法。

第一节 主题标引课程的意义与任务

一、主题标引课程的意义与特点

生物医学文献信息资源在整个医学信息资源中占有极其重要的地位。组织文献信息资源,建立手工检索系统或计算机检索系统,首先要做的工作是需要对大量的文献资料进行加工处理,只有把大量的文献按一定的体系和规则加以处理后,才能将它们有效地建成情报检索系统,进行情报检索。在医学信息本科专业开设医学文献主题标引课程,让医学信息学专业的本科生掌握生物医学文献主题标引的规则与技巧,为我国医学信息事业培养专业人才,是医学信息学教育的重要任务之一,有着十分重要的意义。

在诸多的文献信息资源的组织方法中,主题法是最重要的组织方法之一。它直接以代表事物的名称和概念的名词术语作为文献主题的标识,并依据这些标识符号的语义和字顺,编排和组织文献信息资源,构成情报检索系统。在信息检索过程,尤其是网络信息资源的查找过程中,主题法以其专指性、直接性和组配灵活的特点,成为主要且广泛使用的检索方法。

二、主题标引课的任务

医学文献主题标引是一门从主题角度深入揭示医学及其相关学科文献内容的专业课程,是情报语言学中叙词语言的一种非常重要的实际应用,其主要任务是使学生在对情报检索语言尤其是叙词法的基本原理与方法全面了解的基础上,掌握生物医学文献主题标引的规则与技巧,达到标引的准确性和一致性。

三、学好本课程几点要求

医学文献主题标引作为一门最能体现医学信息学专业特色的核心课程,是医学信息学专业课程中较难掌握的课程之一,如何学好本课程应注意几点:

1. 必须具备一定的医学及其相关学科专业知识

医学文献记载的都是医学及其相关学科的专业内容,文献中所论述的对象和观点,都是生物医学科研工作者和临床医生很专业的课题,如果不掌握一定的医学专业知识,不了解该课题的价值及创新点,在标引过程中就无法进行主题分析和准确提炼主题,导致标引误差。因此,掌握医学信息学专业的医学课程,是学好本课程的基础。

2. 应熟练掌握情报学专业知识和专业英语

学好本课程必须熟悉检索语言的基本理论、结构和使用规则,尤其是需要掌握主题语言中的叙词法的结构、原理与概念表达方式。此外,了解情报工作的各个环节尤其是情报检索过程和熟悉情报用户的需求,对于完成标引工作是有益的。本书针对大多数医学信息专业未专门开设情报检索语言课程这一情况,就情报检索语言的基本理论、结构和使用规则做了介绍,为掌握医学文献主题标引规则与方法打下基础。

医学文献主题标引规则与方法,主要是以美国国立医学图书馆(NLM)创建的大型生物医学文献集成系统——医学文献分析与检索系统(MEDLARS)的标引方法与技术为基础。《医学主题词注释字顺表(MeSHAAL)》及其标引规则年年变更,要动态性掌握这种变化,学好本课程应具有较高的英语水平。

3. 加强动手能力的培养

医学文献主题标引规则较为复杂,为了便于学习,在阐述每一项标引规则时,均援引具体实例。标引工作是一种技巧和艺术,要掌握它,不是一朝一夕能做到的,必须经过大量的标引实践,在“动手”过程中,积累标引经验,从而掌握与理解标引规则。

第二节 情报检索语言概述

一、情报检索语言概念

我国著名情报学家张琪玉教授对情报检索语言的定义是“根据情报检索的需要而创制的人工语言”。苏联情报学家切尔内对情报检索语言的阐述是:“一种专门的人工语言,它用于描述(表示)文献的中心问题或主题和外表特征,以便以后在其他文献集合中查找需要的文献,或者用于表示情报提问的内容和检索需要的文献。”作为一种专门的人工语言,它的作用是在标引过程中用来描述文献的内容特征(或外表特征),从而形成检索标识;在检索过程中用来描述检索提问,从而形成提问标识。这种检索标识(提问标识)需要用语言形式来表示,这种特殊的语言就称为情报检索语言,简称为检索语言或标引语言。

二、情报检索是情报工作中的重要环节

所谓情报工作,就是运用科学的方法,把有用的科技信息、知识和新的科研成果,经过搜集、分析、综合加工、存贮和检索,有计划、有目的、准确、及时地提供给使用者。在现代信息社会,科学技术文献信息数量庞大,内容专深且互相渗透,如何组织文献信息资源,即按照一定的科学原理、规则和方法使大量的、无序的信息资源成为有序的信息资源,满足社会对情报的需求,解决信息资源需求与利用之间的矛盾,有效地进行信息资源管理和实现信息资源增值,是每个信息工作者所面临的基本任务。情报检索即情报的存贮和检索,在整个情报工作过程中占有十分重要的地位。

三、情报检索语言与情报检索的关系

情报检索的全过程包括两个基本环节:情报的存贮和检索。情报存贮是指要把搜集的情报源,即文献、数据等按一定的体系和规则进行处理和存贮,从而编制检索工具和建立检索系统;情报检索即是利用这些检索工具和检索系统来查找所需的情报。所以,情报的存贮和检索是两个有着密切联系的过程,存贮情报时,文献标引人员首先要对各种文献进行主题分析,即把它所包含的情报内容分析出来,使之形成若干能代表文献主题的概念,并用情报检索语言的语词把这些概念标示出来,然后纳入检索工具或检索系统。当检索情报时,情报检索人员也首先要对检索课题进行主题分析,即把它所涉及的检索范围明确起来,使之形成若干能代表情报需要的概念,并把这些概念转换成情报检索语言的语词,然后从检索工具或检索系统中查找正是用该语词标引的文献,从而找到包含有所需情报的文献。由此可见,情报检索语言在情报检索系统中起着语言保障的作用,是沟通标引人员和检索人员双方思想的桥梁。如果没有情报检索语言作为标引人员和检索人员的共同语言,就很难使标引人员对文献情报内容的表达和检索人员对相同内容的情报需要的表达取得一致,情报检索也就不可能实现。

四、情报检索语言的特点

目前世界上有各种类型的情报检索语言,作为一种人工语言,它可以是从自然语言中精选出来并加以规范化的一套词汇,也可以是代表某一类事物的某一特征的一套代码或代表某种分类体系的一套分类号。既然是人工语言,它就和人们日常科技交往所使用的自然语言是有所区别的。其特点是:①必须具有较小的容量,能简单明白又比较专指地表达文献及检索课题的主题概念,容易将其进行系统排列;②必须具有准确的语义概念,检索标识与概念一一对应,排除了多词一义、一词多义和词义含糊的现象,并且还能显示出概念之间的相互关系;③需要编制一种类似词典那样的工具——检索词典,情报检索语言由词汇和语法组成。词汇是指收录在分类表、词表中的全部标识,一个标识就是它的一个语词,而分类表、词表则是它的词典;语法是指如何创造和运用那些标识来正确表达文献内容和情报需要,以有效地实现情报检索的一整套规则。

五、情报检索语言的发展

情报检索语言发展至今,主要分为分类语言和主题语言两大类型。曾在很长一段时间内,分类语言居于垄断地位,人们习惯于将各种概念按学科性质进行分类和系统排列,即用分类号表达和描述文献内容的方法来组织文献资料。分类语言包括等级体系分类语言(体系分类法)和分析—综合分类语言(组配分类法)。体系分类法主要应用概念划分与概括的方法,组配分类法主要应用概念分析与综合的方法。组配分类法是体系分类法的发展,它创建的目的,是为了克服体系分类法不能无限容纳概念的局限性,以及它的类目的单线排列方式所造成的“集中与分散”的矛盾(事实上,体系分类法采用的某些方法,也含有概念分析与综合的方法,实行了某种意义上的组配)。然而分类法由于受到等级划分的限制,这种类目层面上的组配,在表达和描述任意专指主题概念时存在着较大的缺陷,加之它在直观性、动态性等方面存在的问题,它的应用受到不同程度的制约。在这种背景下,主题语言应运而生。所谓主题语言,是指用自然语言中经过规范化或未经规范的语词来表达各种概念,以概念之间的形式逻辑作为语法和构词法,用一组词语作为检索标识的检索语言。各种概念不管其相互关系按字顺排列,概念性、控制性、组配性、语义性和动态性是其基本特征。主题语言包括标题词语言、单元词语言、叙词语言和关键词语言等。主题语言发展至今,其理论、方法不断完善、更新,自身类型也不断演化、丰富,并经历了从手工标引和检索到计算机标引和检索的发展过程。目前分类法语言和主题法语言也互相渗透,各种方法互相采用。随着情报技术的不断发展,情报检索语言的发展和研究热点主要表现如下几个方面:①情报检索语言的标准化。在情报技术发展过程中,相继出现了各种情报检索语言标准。比如我国编制的《中国图书馆分类法》、《汉语主题词表》等,国

外的检索语言工具有《医学主题词注释字顺表》、《国际十进分类法》等。这些工具规范了人们的情报检索工作,但如何将众多的标准有机结合起来,进一步提高情报检索语言的标准化,是情报研究的一大任务。

②情报检索语言的一体化。情报检索语言的一体化是指分类法主题法的一体化,包括编制一体化词表和编制两种语言的对应表。③自然语言与自动化检索系统研究等。

除了上述两大类型外,情报检索语言还有代码语言和引证关系追溯法等类型。

第三节 情报检索语言构建基础

要掌握情报检索语言构成原理、词表结构和表达文献主题概念及其相互关系的作用,就需要有概念逻辑、知识分类、术语学等方面的基本知识。

一、概念、概念逻辑

(一) 概念

概念是事物本质属性的概括,以语词作为表现形式。它是人类思维的细胞,构成判断和推理的要素,形成各种思想和知识的前提;也是科学认识的总结,表达交流思想知识的工具。

概念都有内涵和外延。概念的内涵是指概念所反映的事物的本质属性的总和,是概念质的表现。概念的外延是指概念所反映的对象的范围,是概念量的表现。内涵有多少的区别,即概念所反映的本质属性有多有少;外延有大小的不同,即概念所指的范围有大有小。在概念的内涵与外延之间存在一种反变关系,即:当内涵增多时,外延就缩小;当内涵减少时,外延就扩大。

概念逻辑是一种科学思维方法,运用它能揭示事物的本质属性及各种事物之间的联系与区别。它包括:明确概念的类型与含义,如集合概念与非集合概念,具体概念、抽象概念与关系概念,单称概念、特称概念与全称概念等;确定概念的内涵和外延,用语词命名表达;确定概念之间的联系与区别,明确概念之间的关系。

(二) 概念间的逻辑关系

事物的概念之间,有的没有关系,有的则有一定的关系。按概念外延是否有相同部分,概念之间的逻辑关系可以归为以下几种类型:

1. 相容关系

相容关系指两个概念的外延部分或全部重合,包括同一关系、属种关系、交叉关系、整体与部分关系、全面与方面关系。

(1) 同一关系,是指两个概念外延完全相同,如“艾滋病”与“获得性免疫缺陷综合征”。

(2) 属种关系,一个概念的外延完全被包含在另一个概念的外延之中。外延大的称为属概念,外延小的称为种概念。概念的属种关系,不仅具有相对性和包含与包含于的区别性,而且还具有系列性和层次性,可以形成概念的等级关系。如“医学”与“临床医学”。

(3) 交叉关系,指两个概念外延部分重合,如“男青年”与“医生”。

(4) 整体与部分,是表示事物整体的上位类与表示事物某一部分的下位类之间的关系,如“脑”与“脑干”。

(5) 全面与方面关系,是表示事物总体的上位类与表示事物某一方面的下位类之间的关系,如“肿瘤”与“肿瘤诊断”。

2. 不相容关系

不相容关系指外延完全不重合的概念之间的关系,包括并列关系、矛盾关系和对立关系。

(1) 并列关系,指一个上位类依据某一分类标准划分出的各个下位类之间的关系,如“肿瘤诊断”与“肿瘤治疗”。

(2) 矛盾关系,指同一属概念下的两个同层次的种概念之间,外延完全不同,且外延之和等于其属概念的外延的一种关系,如“奇数”与“偶数”。

(3) 对立关系,指同一属概念下的两个同层次的种概念之间,外延完全不同,且外延之和小于其属概念的外延的一种关系,如“高血压”与“低血压”。

概念之间的逻辑关系是主题语言显示标识之间关系的主要逻辑学基础。

(三) 概念划分、限制与概括

1. 概念划分

划分是明确概念外延的逻辑方法。它是把概念所反映的全部对象按一定标准分为若干种,从而将一个属概念分为若干种概念的过程。划分由母项(被划分的属概念)、子项(划分所得的子概念)、划分标准(划分依据的事物属性)三部分组成。为了正确地划分,划分时必须遵守四条规则:

- (1) 在同一次划分过程中应按照同一标准进行划分;
- (2) 划分所得的子项应该是不相容的;
- (3) 划分必须按照属种关系的逻辑层次逐级进行;
- (4) 划分所得的子项外延之和必须与母项的外延相等。

2. 概念限制

限制是缩小概念外延的逻辑方法,它通过不断增加概念的内涵来缩小概念的外延,由属概念过渡到种概念。

3. 概念概括

概括是扩大概念外延的逻辑方法,它是限制与划分的逆过程,由外延较窄的概念过渡到外延较宽的概念,由种概念过渡到属概念。

限制与划分都是增加概念属性,缩小概念外延的逻辑方法。但是,限制只是从属概念中缩小出一个种概念;划分则是从一个属概念中分出若干个种概念。

情报检索语言中的概念划分、限制与概括,是利用概念内涵由反映事物本质属性的概念因素构成,概念因素的增加或减少可以形成新的概念,概念内涵与外延成反变关系等逻辑特征,对概念进行划分与概括,以便将泛指概念缩小为专指概念,或者将专指概念扩大为泛指概念;同时利用划分与概括中所产生的概念隶属关系和并列关系,形成概念等级体系,作为情报检索语言等级结构的逻辑基础。

(四) 概念分析与综合

我们已经知道具有交叉关系的两个概念的外延重合部分可以形成一个新概念,它的内涵等于原来两个概念内涵之和,与原来的两个概念具有属种关系。由此可知,概念是由概念因素组成的,复杂的概念可以分解为几个简单概念,几个简单概念也可综合成一个复杂概念。情报检索语言就是应用这个原理,将两个或多个内涵较少的概念,用交叉和限制的方法综合为一个内涵较多的概念,形成供组配使用的概念体系,作为组配方法的逻辑学基础。

要建立反映事物的概念之间的正确关系,仅依据概念的逻辑理论是不够的,还必须利用知识分类的成果。

二、知识分类

(一) 知识分类定义与原则

1. 知识分类定义

知识分类是对事物做系统研究的重要方法,是从本质上、原理上揭示事物之间区别和联系的重要手段,是对知识进行系统组织的科学方法,也是利用知识的有效途径。

2. 知识分类原则

知识分类的两条基本原则是客观性和发展性。客观性原则要求知识分类时依据知识对象固有的、客

观存在的区别和联系进行知识的划分和组织。发展性原则要求知识划分和组织时要体现知识从低级到高级、从简单到复杂的发展规律。

(二) 知识分类在情报检索语言中的应用

1. 知识分类是情报检索语言的基础

情报检索语言的基本功能是将文献及文献中记录的信息、知识系统化,以方便人们获取和利用。知识分类已经对知识进行了科学的系统组织,建立了事物分类体系,而且人们通常从知识分类的角度对信息、知识、文献提出要求,所以情报检索语言的概念分类体系应尽可能以知识分类体系为基础。

实际上,分类语言就以知识分类体系为基础建立类目体系,主题语言根据知识分类建立参照系统、编制范畴索引、词族索引,代码语言的代码设计也体现知识分类的成果。

2. 知识分类体系与概念分类体系的差异

尽管情报检索语言的概念分类体系应尽可能以知识分类体系为基础,但情报检索语言的概念分类体系与知识分类体系还是存在着差异的。主要表现在:

(1) 分类对象不同,知识分类的对象是纯粹的知识内容;概念分类体系不仅要文献中的各门科学知识分类,还要对非知识的内容进行分类。

(2) 分类标准不同,知识分类的分类标准只是知识内容的对象属性;概念分类体系不仅要知识内容的对象属性作为主要分类标准,还要以内容的表现形式(即文献外表特征)作为辅助分类标准。

(3) 分类目的不同,知识分类的目的是对知识进行划分和组织,确定各门科学知识在科学整体中的位置,揭示整个科学的内部结构;概念分类体系不仅要文献的内容系统化,还要将文献本身系统化,以便人们查询利用。

(4) 分类体系的结构和表现形式不同,知识分类是一种动态的网状结构,可采用多种描述方式,如线性序列、列表方式、图示方式等,不需要采用号码标记等;概念分类体系一般采用线性结构,用号码标记类目及类目之间的关系,要求相对稳定。

因此,情报检索语言既要知识分类为基础,又不能局限于知识分类;既尽可能利用知识分类的成果,又难以与知识分类体系保持完全一致。

三、术语学基础

主题检索语言中的标识基本上是经过规范化了的名词术语。何谓术语呢?术语是以单义性的方式为某概念命名的语言单位,它是概念和名称的统一体。某一领域的术语集合就是通常所说的专业词汇。术语与一般语词不同,它是在明确了某个概念之后,才考虑用哪个语言形式表达概念,概念先于名称。对于术语而言,总是要求一个概念与一个而且仅与一个名称相对应,一个名称与一个而且仅与一个概念相对应,而且,一个术语所表达的概念,必须依据它在该领域的概念系统中所处的地位而获得定义。此外,为方便交流,术语还具有语言形式简洁、一经构成就相对稳定、多为名词和名词性词组等特点。术语所具有的上述特点与情报检索语言词汇控制的目标非常一致,因此,术语必然成为情报检索语言优先、直接、主要的词汇来源。以概念和概念体系为核心的术语学,也就成为检索语言的基础。

本章小结

医学文献主题标引是医学信息学本科专业一门重要的专业课程,本章节主要介绍本课程的意义与任务,情报检索语言基本概念和构建基础。

关键词

情报检索语言 概念逻辑 知识分类

思考题

1. 情报检索语言体系构建基础是什么?
2. 如何学好本课程?

推荐阅读

- [1] 肖晓旦主编. 生物医学文献主题标引. 长沙:湖南科学技术出版社,2005.
- [2] 张琪玉主编. 情报检索语言实用教程. 武汉:武汉大学出版社,2004.

情报检索语言的类型

内容提要

本章通过对情报检索语言的类型的介绍,详述了五种情报检索语言,即等级体系型分类语言、组配分类语言、标题词语言、单元词语言和叙词语言。并分别对这五种语言的构成原理、规范程度、词表体系结构及使用特点等方面进行了逐一阐述,尤其突出了等级体系分类法和叙词法的描述。

情报检索语言作为一种专门的人工语言,有许多种类型,但任何一种类型必须具有词汇和语法手段,能准确有效地用于情报标引和检索,真正起到情报检索的语言保证作用。

第一节 情报检索语言的种类

我们通常把一部分类法、一部主题词表称为一种情报检索语言,目前世界上的情报检索语言已有数千种之多。情报检索语言根据不同的需要,有不同的划分方法。

一、按规范化程度划分

1. 非规范化语言

也称非受控语言,就是直接从原始信息中抽取出自由词作为检索点的检索语言。该语言对主题概念中的同义词、多义词等不加处理,取其自然状态,因此也称自然语言。自然语言具有不编制词表、选词灵活多变、标引和检索速度快、及时反映事物发展变化、准确表达新概念等优点,在计算机技术飞速发展的时代,自然语言更能满足用户对信息检索时效的要求。但自然语言的缺陷也是十分明显的,即用户表达信息需求的用词不一致将导致他们的检索结果有很大的差别。主题语言中的关键词属于这一类语言。

2. 规范化语言

又叫受控语言或人工语言,就是人为的对标引词和检索词加以控制和规范,使每个检索词只能表达一个概念。这些语言经过规范化控制,检索词和事物之间具有一一对应的关系,从而排除了自然语言中同义词、多义词、同形异义词现象。分类法、标题词、单元词和叙词就属于这一类语言。

二、按组配方式划分

有先组式检索语言与后组式检索语言两种。在先组式检索语言中,又有定组与散组式检索语言两种。

1. 先组式检索语言

意指表达文献内容的检索词在查找文献前已经组配好的检索语言。它有两个分支:其一是定组式检索语言,意指表达文献内容的检索词在编制检索系统时已经组配好的检索语言。标题法和体系分类法属

于这一类型。其二是散组式检索语言,意指表达文献内容的检索词在检索系统中不预先组配好,而只是在进行文献标引时才加以组配的检索语言。叙词法属于这一类型。

2. 后组式检索语言

意指表达文献内容的检索词在查找文献时才进行组配的检索语言。单元词、关键词和组配分类法属于这一类型。

三、按构成原理划分

1. 分类法语言

分类法是将表示各种知识领域(学科及其研究问题)的类目按知识分类原理进行系统排列并以代表类目的数字、字母符号(分类号)作为文献概念标识的一类情报检索语言。分类法的主要特点是按学科、专业集中文献,并从知识分类角度揭示各类文献在内容上的区别和联系,提供从知识分类检索文献的途径。分类法有体系分类法和组配分类法两种类型。体系分类法依据的是“概念划分与概括”原理,如《中国图书馆分类法》;组配分类法依据“概念分析与综合”原理,如《冒号分类法》。不管是体系分类法还是组配分类法,它们都是按照知识组织的原则构建的,即用类目名称(文献概念)表示知识因子,用等级体系以及参照、注释、互见、交替等方法来直接显示类目之间的知识关联,用学科区分及分类号的层次序列作为组织知识集合的基本方式。

2. 主题法语言

主题法是以自然语言的语词经过规范处理后直接作为文献主题标识,并按字顺排列,结合参照体系和其他方法来间接地显示概念之间的关系,提供从事物名称检索文献的途径。主题法的主要特点是按事物集中文献,用参照系统等方法间接显示概念或事物之间的关系,提供事物名称的字顺检索途径。它包括标题法、单元词法、叙词法、关键词法。主题法用主题词表示知识因子,用参照系统来表示知识关联,用语词标识系统作为组织知识的基本方式。

3. 代码检索语言

代码检索语言是指对事物的某方面特征,用某种代码系统来表示和排列事物概念,从而提供检索的检索语言。例如,根据化合物的分子式这种代码语言,可以构成分子式索引系统,允许用户从分子式出发,检索相应的化合物及其相关的文献信息。

4. 引证关系追溯法(引文索引法)

引证关系追溯法是采用引文索引语言来标引文献和建立检索系统的一套技术方法。最早出现在 19 世纪 70 年代,后经美国情报学家尤金·加菲尔德等人加以发展和完善,于 20 世纪 60 年代成功地运用于科技文献的标引和检索,成为国际上颇有影响和吸引力的一种新型索引方法。它是利用文献中普遍存在的参考引证现象,通过两篇或多篇文献来源之间由一篇或多篇文献的相同引文所形成的关系网,为人们提供的一种新的文献归类方法和检索方法。

本章重点阐述分类检索语言和主题检索语言。

第二节 等级体系分类语言

1876 年杜威创立《杜威十进分类法》(Dewey Decimal Classification, DDC),成为图书分类法发展史上的里程碑,它标志着杜威时期的开始,开辟了图书分类法的新纪元。1901 年,美国《国会图书馆图书分类法》(Library of Congress Classification, LCC)的问世,标志着 20 世纪大型体系分类法技术已经成熟,并得到广泛应用。我国的《中国图书馆分类法》、《中国科学院图书馆图书分类法》和《中国人民大学图书馆图书分类法》是国内影响较大的三部综合性分类法。

一、构成原理

等级体系型分类语言也称体系分类法,是一种将所有的类目组织成一个等级系统,并且采用尽量列举的方式编制的分类法,故又称作列举式分类法、枚举式分类法。这种分类法通常将类目体系组织成一个树状结构,按照层次划分,逐级列出详尽的子目,并在以线性形式显示时,以缩格表示类目的等级关系。由于这种分类法通常是依据传统的知识分类体系编制的,人们习惯上也将其称为体系分类法。体系分类法适合于从学科、专业系统检索和利用文献,具有列举式类目、分类符号和标识、等级制结构、直线性序列等特点,是目前使用较多的一种分类语言。

体系分类法的特点是:①从一定的角度出发有层次地揭示信息资源,类目展开比较系统;②分类结构显示直观,易于把握、便于使用;③标记简明,号码单纯,适于分类排架,也可以用于组织分类检索工具。其不足是:①揭示专门主题能力差,因为体系分类法的类表是列举式的,而要详尽无遗地列举一切事物概念,在事实上是不可能的;②类表具有一定的凝固性,不便于根据需要随时改变、调整检索途径,不能进行多角度检索;③它的类目体系是单线排列的,这也就限制了使一个复杂的文献主题概念得到多向成类的可能性,无法根据现代科学的发展自动生成新类,难以与科学的发展保持同步;④大型列举类表一般类目详尽、篇幅较大,对类表管理的要求较高。

体系分类法提供的是分类检索途径,它的检索标志就是给定的分类号码。用这种检索途径的前提是,必须确定检索课题所属的学科类目,找出该类目的分类号。

二、类目的划分

类目的划分,是指依据一定的属性或特征对类目的外延进行区分,生成一组子目。类目划分是类目体系建立的基本方法,是逻辑方法在信息资源组织中应用的重点,类是分类体系的基本构成单元。

在类目体系展开过程中,一般须遵守一定的逻辑划分规则:其一,每次划分只使用一个标准,一般不得同时采用两个或两个以上的标准,如,当教育文献在依据教育等级进行区分时,就不能同时使用涉及的国家或教育活动的具体内容作为标准进行区分;其次,划分应该穷尽被区分类的外延,使划分后子类之和与母类的外延相等,使该分类领域下所有的内容对象有类可归;其三,划分后各子类应相互排斥,界限分明,类目之间不应存在相互交叉现象。实际上,只要严格贯彻第一条规则,划分出来的子类必然会相互排斥。

三、类目的排列

类目排列有自然序列和人为序列两种。在一般情况下,优先采用自然序列。

自然序列是指按自然界的发展演化进程排列,按社会及各种社会现象的发展过程排列,按概念的系统性排列,按从低级事物到高级事物(或从简单事物到复杂事物)的次序排列,按物质结构的某种次序排列,按生产流程的次序排列,按习惯次序排列,等等。

人为序列是指从重要事物类目到次要事物类目的排列次序,从有现实意义的事物类目到仅有历史意义的事物类目的排列次序,从有较多文献论述的事物类目到只有较少文献论述的事物类目的排列次序,从抽象的、理论的、一般性问题的类目到具体的、应用的、个别性问题的类目的排列次序,等等。

四、类目名称及含义

类目名称是等级体系分类语言的“语词”,它限定着所表达的事物概念的内涵和外延。对类目名称的要求是确切、科学、简洁。一个类目可以表达一种事物,或几种事物,或一类事物,或一种事物的一个方面,或几种事物的同一方面,或一类事物的同一方面,等等。

类名字面意思在大多数情况下不能直接表达事物概念,其含义要受上位类限定,有时还要受类目注