

文献学与排序论

情报检索理论和实践教程

〔德〕W. 高斯 著



科学技术文献出版社

C-35C
F

文献学与排序论

情报检索理论和实践教程

[德] W. 高斯 著

史秀英 于良娇 译

汤兆魁 校

DZ 59/19

科学技术文献出版社

内 容 简 介

本书共分36章,详细阐述了文献加工的基本概念、步骤和方法,并介绍了情报服务、情报检索和计算机应用的基础知识。各章还附有思考题和参考答案。可作为情报、图书、档案等部门进行业务培训和大专院校情报、图书、档案等专业师生的教材;同时也是情报、图书、档案、计算机应用、管理等部门工作人员的有益参考书。

DOKUMENTATIONS- UND ORDNUNGSLEHRE

Wilhelm Gaus

Springer-Verlag Berlin Heidelberg New York Tokyo

1983

文献学与排序论

〔德〕W. 高斯 著

史秀英等 译 汤兆魁 校

科学技术文献出版社出版

海洋出版社印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

850×1168毫米 32开本 16.75印张 432千字 插页1张

1989年11月北京第一版第一次印刷

印数: 1—3000册

社科新书目: 224—156

ISBN 7-5023-0829-6/Z · 119

定 价: 9.95元

译者的话

随着科学技术的迅速发展，计算机技术广泛地应用到各个领域。在科技情报界，为了能够向广大用户全面、准确、迅速地提供文献情报，西方国家在60年代末、70年代初纷纷建立各种规模的科技情报数据库，与此同时形成了一个新的独立学科——文献学，亦称文献加工。在德意志联邦共和国规定，大学毕业之后必须再专修一年文献学才能取得文献员资格。

文献加工包括对文献的采集登录、标引、存贮和检索全过程，是文献服务工作的现代模式。文献加工的质量直接影响到向用户提供情报的准确性。如果使用计算机建设数据库，那么数据库的生命将取决于文献加工的质量。

《文献学与排序论》这本书用大量实例详细地阐述了文献加工的重要性及其具体方法。它内容新颖、条理清楚、通俗易懂。除文献加工的一般方法和问题外，本书还介绍了有关情报工作的基本概念、情报检索、情报服务以及计算机在情报检索中的应用等方面内容。这些正是目前从事情报、图书和档案工作的人员所必须学习和掌握的。

为了节省篇幅，我们翻译中删去了在其它出版物中介绍过的关于穿孔卡片和比孔卡片等章节（即原著第16~19章）。

由于我们各方面的水平有限，译文难免有欠缺和错误之处，欢迎读者批评指正。

译者

1988年12月

前 言

乌尔姆大学医学文献加工学校，每年在主要学科“医学”、“文献加工”、“数学和统计”、“数据处理”以及“组织学”等方面，培养36名毕业生成为医学文献工作人员。自1969年这个学校成立以来，本人就在这里教授“文献学与排序论”。本书即我的教学笔记。这门课程需要120个课堂学时和40个练习学时，可以在第一年开设，不需要其它基础知识。

文献加工的中心问题是情报检索。本书材料涉及了其它一些文献学科，如：

- 图书馆学
- 医学文献加工系统
- 医学文献加工
- 复制方法和缩微
- 法学和户籍管理学

无疑，本书是针对医学编辑的，很多例子取之于医学文献。但是它适用于各领域的文献工作者，而且不需要具备医学知识。

副标题用“教程”一词是想说明，本书不强调全面性和科学的彻底性，而是一种教学和实际应用两方面的融合，并留有余地，供读者思考。章节按认识的发展安排，较少地考虑事物的类别关系。有些课程是为概括了解、加深认识或复习安排的。就连六种排序原则也是主要从实际应用方面确定的，而比较少地从纯科学理论出发。为了使教程更加直观和便于讨论，本书收进了一些实习用的词表和资料。每章后面附的习题是测试题。供自学读者练习和自我测试用。

威廉·高斯

1983年7月于乌尔姆

排字原则
分类
索引
组面
概念组配
语义分析
目由同

L 8
L13
L15
L16
L17
L27

排序系统
等级分类—字顺
结构编制
补充叙词
修订

文献单元	L33
采集形式著录	L5

索引
文摘
L6
主动文献服务L29

叙同档 文档和情报档	L 4
D·D·矩阵 叙同档 联机检索	L23 L26 L34

检索质量 查找策略 L24 L25
联机检索 L34

用户
需求 L31

文献单元流程方向

得

23

例

必要性和可能性

概念一名称
术语控制
数据文献加工

工序·经费
文献档
计算机使用

建筑物主题词表 L20
现有词表 L21

目 次

第1章	文献加工的目的和基本特征	(1)
第2章	情报流—文献加工的必要性	(10)
第3章	情报的概念和性质	(16)
第4章	文档和情报存贮器	(22)
第5章	外表特征描述和内容揭示	(40)
第6章	文摘	(48)
第7章	概念和名称	(53)
第8章	分类排序原则	(64)
第9章	概念等级结构	(72)
第10章	按字顺排序和分类排序	(85)
第11章	诊断码: 国际疾病分类法和临床诊断码	(92)
第12章	十进分类法	(104)
第13章	索引排序原则	(121)
第14章	按索引排序原则进行文献加工的示例	(130)
第15章	组面分类排序原则	(148)
第16章	概念组配排序原则	(160)
第17章	语义分析排序原则	(165)
第18章	词表(排序系统) 的结构	(175)
第19章	词表的编制	(191)
第20章	建筑物叙词表示例	(205)
第21章	关于现有词表的讨论	(276)
第22章	连号、职号、时间描述符和语种描述符	(302)
第23章	文献叙词矩阵	(314)
第24章	查准率和查全率	(327)
第25章	检索和查找策略	(341)

第26章	叙词存贮器和检索辅助工具·····	(354)
第27章	自由词“检索”排序原则·····	(367)
第28章	术语控制·····	(381)
第29章	主动的文献服务·····	(391)
第30章	数据文献的特点·····	(398)
第31章	斟酌用户的需求·····	(408)
第32章	词表的修订·····	(415)
第33章	工序和经费·····	(427)
第34章	联机检索·····	(440)
第35章	在文献加工中使用计算机·····	(449)
第36章	概述和标识符号·····	(456)
附录	·····	(464)
	习题答案·····	(464)
	引用文献·····	(516)

第1章 文献加工的目的和 基本特征

1.1 文献加工的目的

本书所论述的文献加工目的和意义在于有针对性地检索并利用文献和情报。不论文献或情报的收集，还是排序本身都不等于文献加工。这里应首先强调文献员的主观努力和认真负责的工作。因此情报文献的收集和排序并不是目的，仅仅是文献工作的一部分，还算不上是完整的文献加工。英文“文献加工”一词确切地说是“情报检索”，按字面直译为“查找情报”。文献加工的质量取决于是否能找到与查找问题相关的、而且是完全相关的文献。

当然，除了情报检索外，还有其它形式的文献加工，比如由国家图书馆、书库、档案馆、公司、政府机构、政党或联合会搞的文献工作，更多地用于情报的收藏、展示或其它用途。但是，这些不属于本书论述范畴。

文献工作者在情报检索工作中不是为自己、而是为用户工作。某一文献的用户可以是科学家、管理人员、大学生、技术人员、医生、商人、政治家、法律工作者等等，简而言之，是所有需要情报的人。归根结底，文献加工的目的是回答用户的提问，所以提供给每个用户的文献和情报都应该是相关的，而且是真正相关的。

1.2 文献单元

对文献进行处理,或者一般所谓的文献单元,即指:一本书、一期杂志、杂志中的一篇论文、书中的一页内容、一句话(如圣经经文、法律条文)、一个说明、一个测试值、一个数据或一个事实(如:乌尔姆大教堂的高度=161.6m,德意志联邦共和国的面积=248 577km²)、一种材料的数据单、画廊的一幅画、劳动局或军队里的一份人事档案、一份病历、一种专营药物、一项医疗诊断(如:脾大、经常咳嗽、体温37.2℃)等等。

根据文献单元的种类,分别称作文献、数据文献、材料文献、图象文献、人事档案文献、病历文献、药剂文献、医疗诊断文献等等。

“文献单元”一词虽然比“文献”一词冗长,但更符合实际情况。“文献”一词还有一种法学意义。在口语中,这个词常指有价值的事物,而不是指日常所见的事物。为了避免混淆,我们应该用“文献单元”。当然,有时为了简便起见,也在等同意义上使用“文献”一词。

很多文献员区别对待“文献单元”和“文献有关单元”。这二者之间的区别就在于有些文献机构并不拥有书籍、论文、病历等等,而只有它们的详细说明,例如:卡片索引。这样,书籍、论文、病历等就是“文献有关单元”,而隶属于它们的索引卡片就是“文献单元”。假如把一个大停车场的车辆,或大工厂里的机器设备比作“文献有关单元”,那么这两者之间的区别就更直观了。即使文献机构拥有文献有关单元,最好也为其各做一张索引卡片或类似的文献单元,用它来代替文献有关单元工作,是一件方便有益的事。从(专门意义上的)文献单元过渡到文献有关单元,或者反过来都不困难,因为文献单元是文献有关单元的代表。所以下面不再区别这二者,而都用“文献单元”一词。

1.3 文献的价值

在决定一份文献是否要进行加工前,首先要进行审查,看它是否值得加工,即这篇文献是否具有文献价值。这要按照具体情况,主要是文献机构的专业范畴和用户范围来决定。决定的关键不是文献员认为该文献好坏,而是看其是否属于该专业范畴,是否与文献加工的主题有关系,用户对其有无兴趣等诸因素。例如,医药公司内部的文献加工部自然会认为,他们生产的全部药品及与之竞争的药品资料都很重要,从而都具备文献价值。文献员自己在审查文献价值时不应成为判断文献好坏的裁判。当然会有一些文献,由于内容关系,明显不符合全体用户的要求,因而被认为没有文献价值。

1.4 文献加工的步骤

文献加工分为四个步骤:

1. 采集和登录

首先要明确文献机构所要加工的文献选题,并将符合这一选题的文献采集进来。重要的是不要遗漏用户认为重要的文献,否则就会造成文献的不完整性。对采集来的文献加以登记(如作者、题目、种类、出版日期及其类似著录项)。

2. 标引

文献内容的揭示叫作标引。可以理解为用所谓叙词(descriptor或index term)确定并标识一份文献的内容。在最简单的情况下,叙词就是主题词。形象地说,每份文献有一个标签,上面有一个或几个叙词,给出了文献的信息。这样,标引也就是对文献的内容揭示,实际就是为文献赋予叙词。示例: 613 号文献

论述骨科疾病（“骨科疾病”就是叙词）。搞标引工作的人就称作标引员。

3. 存贮

文献加工的第三个步骤是存贮。要区别文献存贮和叙词存贮。文献存贮是指文献以其完整的形式典藏保管，如文件集、档案、书库等。目录也属于这一类。如果要区别文献单元和文献有关单元的话，那么书库即是文献有关单元档，而目录、馆藏目录或文献在书架上的定位索取号索引（与书库藏书排序一致）就是文献单元档（狭义）。根据排序论观点，书库与其目录的结构一样，所以都称作文献档。

叙词档中只存贮标引文献的叙词和记录这些文献的索取号，一般不存贮文献单元。查找时只用叙词。为了便于查找，叙词存贮采用另一种结构。叙词档最简单的形式是主题词索引。

4. 检索

检索是有目的地查询并找出有关某一主题的文献单元。所谓有关主题叫作一般提问或口头提问。但要进行检索时必须用叙词组成检索提问式。叙词档根据检索提问式有目的地询问相关文献。叙词档的询问结果是文献索取号。根据索取号可以从文献档中取出所要文献。从事检索工作的人叫检索员。

如上所述，文献加工的首要任务就是要满足用户所提的一切查询需求，而且要尽可能完全准确地提供他们所需要的文献。因此，文献的标引和存贮只是检索的前处理工作。显然，文献机构用于标引和存贮的工作量远远超过检索所花的工作量。

1.5 排序系统（词表）

在标引时，人们或者根据科学的专门术语凭直觉自由地选择

叙词（自由标引），或者从一部事先统一规定的词表中选择恰当的叙词（受控标引）。在受控标引中，把备选的全部叙词及其说明，注释的总和称为排序系统（词表）。它保证了标引和检索用词的同一性。使用统一的受控词汇，即叙词，对文献加工来说非常必要。它能够保证那些针对提问存贮的文献都能检索出来。而用自由词标引则可能出现这种现象：用“汽车”这个词标引某文献单元，而检索时却用“小轿车”或缩写“Pkw”提问，至使这一文献单元不能命中，尽管它对该提问是相关的。

文献机构在搞文献受控标引之前，必须具备排序系统（词表）。当然也可以采用其它单位或自己编制的词表。词表应该有专指性和针对性（如：汽车制造业词表、塑料化学词表、临床诊断词表等等）。词表的大小取决于它所包含的叙词量的多少。最简单的词表只是一份受控叙词的一览表。如果一部词表除了叙词外还包括附加的语义相关词、很多注释及说明，那么就可以称其为扩充很好的大型词表。

每一部词表都是根据一定的排序原则编制的。本书将专用一章来论述六种不同的排序原则。

1.6 索取号和词号

为了方便工作，赋予每份文献单元一个标识符，通常是一个号码。这个简短单义的标识符叫作索取号（Signatur）。索取号说明文献单元在文献档中的位置。同样，也可以赋予叙词一个单义标识符或缩写符号，叫作词号（Notation）。词号标志一种主题及其在叙词库中的逻辑位置。索取号和词号要具备下述特性：

——单义。即每个索取号只代表一份唯一的文献单元，而每份文献单元也只有一个索取号。同样，每个叙词只有一个唯一的词号，每个词号也只代表一个唯一的叙词，用数学语言讲，索取号是文献单元的对应表示，而词号是叙词的对应表示。

——简洁、便于使用。

——便于记忆。应该易读、易懂、易于理解和记忆。

——可以分类。所有的索取号（或词号）都应该能按一定的顺序排列。

——适于机器管理。只局限于打字机或电子计算机输出设备字符组范围内，而上标、下标都不适于机器管理。

——可扩充。便于收入新的文献单元和新的叙词。

这样，索取号是文献单元的简洁而明确的标识符；词号则是叙词的简洁而明确的标识符。联系起来看，词号标识了文献单元的内容，因为词号是叙词的代码。索取号和词号示例：613、A23、64.3、C-F413、82-513。

文献机构在选择索取号和词号时一定要注意它们的外形区别，例如索取号只用纯数字（数字加逗号、点等），而词号则以字母打头。一般情况下索取号是由年号和年度内的流水号组成的，如：82-1，82-2，82-3……。

1.7 工序

对一份新的文献进行加工一般有下列工序：

——采集文献单元；

——审查文献的价值；

——审查是否已有相同的文献（查重）；

——登记并赋予索取号；

——标引；

——将所标叙词连同索取号一起存入叙词档；

——将文献单元存入文献档。

检索一般需有下列工序：

——通过谈话、问答等方式了解用户问题；

——编写检索提问式；

- 查询叙词档，查询结果是相关的文献索取号；
- 按索取号从文献档中取出相应的文献；
- 查看检索结果，如果不满意，则另换检索提问式全部或部分地重复上述工序；
- 将检索结果交给用户。

这是大多数文献机构的典型工序。根据机构大小、所用词表的不同及地方条件等情况可对工序进行增减。详细情况将在第33章《工序和经费》中论述。

1.8 文献的典型结构

大部分文献都用词表。词表是按照一定的排序原则编制的。此外，文献加工还包括文献档和叙词档。新收录的文献要进行标引，为此要用词表。标引过的叙词存在叙词档内，文献存在文献档内。用户提出问题，然后必须将这个问题转换成检索提问式。这里也要用到词表。把检索提问式输入叙词档，检索出索取号，再根据索取号到文献档查找。把检出的文献（或复制件）作为检索结果送到用户手里（见图1.1）。

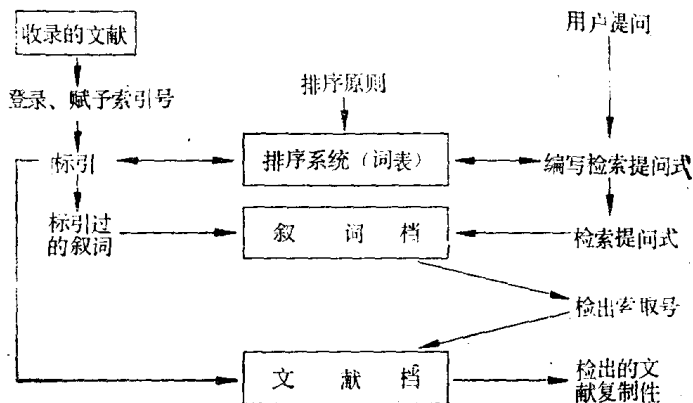


图1.1 受控标引文献的简化结构

加框的项目是物理的（不只是逻辑的）；箭头表示项目之间的关系。

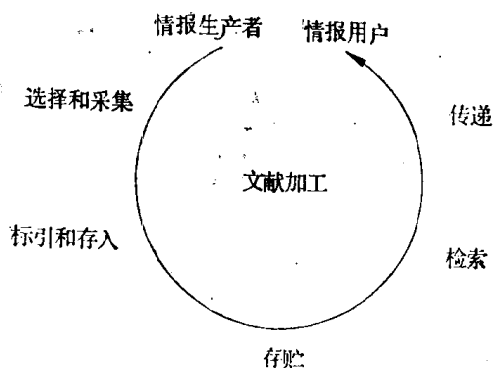


图1.2 文献加工作为情报生产者和情报用户之间的媒介

文献加工的任务是传递情报。这样，文献加工就成为作者、数据提供者、情报生产者与需要情报的用户之间的连接枢纽（见图1.2）。情报生产者和情报用户之间的区别只是形式上的，事实上同一批人、同一个研究所或公司往往既是情报生产者，也是情报用户。

1.9 文献加工示例（A）

某人收集约100份旅游说明书，每年假期旅游一次。每次旅游路线都从说明书中选出。那么，他应该怎样借助词表、叙词档、文献档、索取号等来对他的旅游说明书进行文献加工呢？

答案：

由于收藏量小（100份文献）、检索次数少（每年一次），不值得进行文献加工。最好是每年旅行前从头翻阅一次，直接决定路线。

1.10 文献加工示例（B）

某人在专业课上详细地做了笔记。他将每页笔记作为一个文

献有关单元，并在考试前做一次文献加工。他的全部课堂笔记就是文献档，页码作为索取号，用一个或几个主题词（叙词）标引每一页（文献单元）。他将所用的主题词按字顺排列成表（叙词档），并注明它们所在的页码。如果在结业考试时允许他带课堂笔记和其它辅助用具，考试又局限于课堂所讲的内容，而且他的文献加工工作也做得很出色，那么，他在考试中一定会得到很高的分数。

1.11 习题

1. 举例说出三个文献有关单元及其所属的文献单元。
2. a) 文献加工有哪四个基本步骤？
b) 根据这些步骤说明情报检索意义上的文献加工与经过精心排序的典藏之间的区别。
3. 说明索取号和词号之间的共性和区别。
4. 某些国家的公民都有所谓“普通公民证件”，由出生年月、性别、是否多胞胎、出生地点及其流水号组成。
a) 从国家管理观点来看，这种公民证件是索取号还是词号？说明理由。
b) 某医院的病历是按公民证件排列的。此时公民证件所起作用相当于索取号还是词号？说明理由。
5. 你如何理解“检索提问式”？提示：解释用自然语言和用系统语言提问之间的区别。
6. 比较“标引一份文献有关单元”和“编写一个“检索提问式”两个步骤。说明这两个步骤之间的共性和区别。
7. 检索中的哪一个工序相当于标引？