第14卷 第1期 西北地震学报 Vol. 14, No. 1
1992年3月 NORTHWESTERN SEISMOLOGICAL JOURNAL March, 1992

中国地震科技文献数据库（英文版）建设的质量控制

高 树 心

（国家地震局兰州地震研究所）

摘 要

本文结合《中国地震科技文献数据库（CSJP）（英文版）》多年的建库实践，论述了在分散标引加工条件下，文献库建设数据质量控制的重要性和复杂性。提出了该文献库建库模式，以及文献数据分散标引加工过程中的质量控制方式、控制目的和控制措施。并强调指出：为保证CSJP（英文版）文献库的质量，建立高度权威和高效管理功能的控制中心是必要性。

引 言

我国近十年来文献库的建设有了较快的发展，但仍处在初级阶段，截止1989年3月的统计，全国77种自建文献库中达到5万条记录的仅一个。总记录量仅110万条。文献库用户不多，大多自建自用。能联机公共服务的不多；文献库进入国际市场或进行交换的也不多。尽管如此，库的质量问题已引起各建库单位的普遍重视。

国家地震局着手建设了《中国地震科技文献数据库》，该库的英文版经过七年的研制与建设已初步建成。本文根据多年的建库实践，论述了该库的数据质量控制问题。

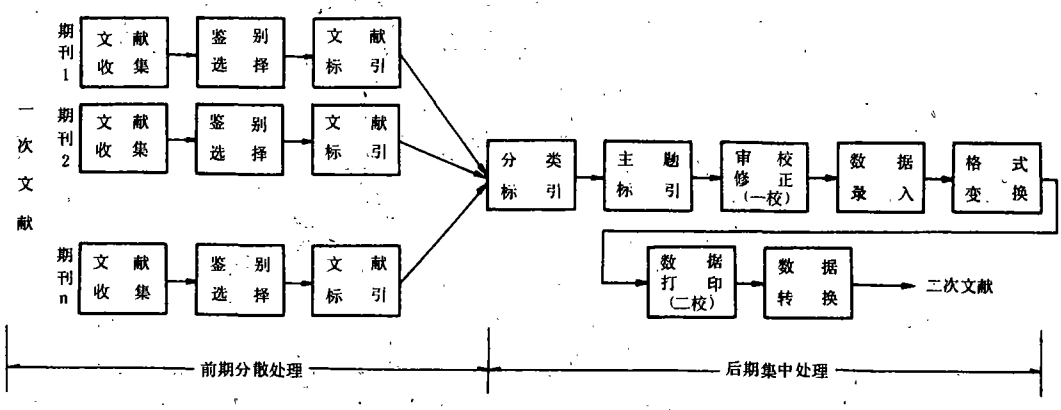
CSJP（英文版）文献库建库模式

计算机情报检索系统是以文献数据库为中心的系统。文献数据库做为存储和检索科技情报的源泉，一是要收集快；二是要数据准。前者反映了建库效率，后者反映了建库质量。一般一个文献库的数据质量包括：数据的准确性；数据的一致性；数据的完整性；数据的稳定性等多方面要求。

CSJP（英文版）文献库是一个文件型数据库。它收集我国出版的六十余种科技期刊中，有

关地震预报研究的学术论文。年文献报导量一千余条。其中二十余种为国家地震局系统正式出版的刊物（以下简称核心期刊）

一般概念认为，对于这样一个小型的专业文献库，生产加工采用集中的方式，质量控制采用人工方式，就可以达到满意的质量要求。但是，由于地震预报是一种多门类多学科的研究领域，完全由一个单位集中完成从一次文献收集到二次文献生成的全过程，不仅在人力、物力上有困难，而且智力也有困难。因此 CSJP（英文版）文献库以“统一规范——分工标引——集中处理”为建库的基本技术路线。1987 年 7 月开始正式建库。1990 年建立了全地震局系统的标引网。核心期刊基本由各编辑部负责标引，非核心期刊由 CSJP（英文版）课题组聘请的人员标引。所生成的工作单寄 CSJP（英文版）课题组集中处理，形成 CSJP（英文版）文献库（图 1）。



CSJP（英文版）文献库加工系统模式

Fig. 1　Processing system mode of CSJP (English edition) documental database

上述文献加工方式，几乎不可避免地引起数据质量问题。如何进行质量控制，是保证文献库能否顺利建成的重要问题，经过 4 年多的建库实践，初步总结出以下一些控制文献库数据质量的体会和措施。

CSJP（英文版）文献库质量控制

1、CSJP（英文版）文献库加工过程控制

集中的加工系统可以及时地对标引人员的工作进行检查，反馈信息和开展标引培训。而分散的加工系统，使得质量控制的及时性、经常性和严格性受到了很大限制。

如图一所示，前期分散处理是分布在不同单位的标引人员，严格按照《CSJP 文献库工作单数据规范》（以下简称“规范”）的要求，标引自己负责的期刊文献，该“规范”是参加建库的文献标引人员，所必须共同遵守的技术文件。

将主题词、分类号的标引放在后期处理之中，由专职人员负责；一校是按“规范”的要求，工作单对照原始文献的审校；二校是将录入的数据打印出来，对照工作单的校对。这些工序与格式变换和数据转换，构成了后期处理的质量控制。

2、CSJP（英文版）文献库的数据类型与相关参照标准

文献库的每条记录由固定的数据字段组成。每个字段有不同的数据类型。明确记录的数据类型是数据质量控制的第一步。本库的数据大致可分以下四种类型。

（1）著录数据——包括篇名、著者、第一著者工作单位、外文刊名、卷期号等字段。著录数据标引的参照标准是一次文献和“规范”。这些数据不需标引人员另行加工，而是要准确无误地反映一次文献的数据信息。

（2）标引数据——包括主题词、自由词、处理码，分类号四个字段。主题词标引、参照标准是《汉语主题词表》，按主题词标引规则，组配规则正确选取主题词。分类号标引，参照标准是《中国图书资料分类法》（现用1985年第二版）。自由词、处理码标引，参照标准是“规范”。这四个字段的标引是加工的重点和难点，也是一校工作的重点。

（3）文摘数据——文摘标引参照标准是“规范”中的有关规定。“规范”的制定参照了ISO214—1976(E)《文献工作——出版物的文摘和文摘工作》。标引的文摘力求用词确切，文句简明扼要。

（4）控制数据——包括控制号和单位代码二个字段。它由编辑人员标引，控制号参照标准是GB3793—83《检索期刊条目著录规则》。单位代码是为数据的一致性设置的。它是数据代换码，不出现在检索命中文献之中。

3、数据质量控制的目标

（1）数据的准确性

就以上四种数据类型而言，数据的准确性是最基本的要求。对于多次重复使用的数据，更应保证绝对准确，否则数据错误将成倍出现。同一卷期中的文献，将相同的字段数据用程序一次性录入时更应注意其准确性。

（2）数据的一致性

数据的一致性应包括横向一致与纵向一致。同一数据无论出现在那篇文献之中，也不论那一时期标引的，都应只有一种数据格式。例如工作单位的英文名称。同一单位，在不同的刊物中应一致，在同一刊物中现在与过去应一致。该系统设计了一个辅助库，单位名称采用了权威性的英文拼写，文献录入时第一著者工作单位仅键入代码，由程序代换给第一著者工作单位字段赋值。

（3）数据的完整性

尽管核心期刊的文献约占年报导量的百分之八十。但从非核心期刊中鉴别、选择适合收录的文献也是一项必不可少的工作。从目前年收录一千余条文献看，该库基本可以反映我国地震科学研究的面貌和学术水平。

（4）数据的稳定性

若词表、分类法、“规范”等发生变动，将导致上述4种数据类型的不稳定，也将影响数据的质量。《汉语主题词表》、《中国图书资料分类法》是国家级文献工作标准。它们的修订再版只是随着科学技术发展作进一步的调整、修改和补充。及时改用最新版本，也不会影响数据的稳定性。

《汉语主题词表》收录专业词汇较粗，专业词表的编制是十分必要的。但必须考虑专业词表和《汉语主题词表》的兼容。所谓兼容，主要指主题词表的结构体系，编辑原则、词汇及

期语义关系等方面力求一致，避免重大矛盾，这样，当改用专业词表时，仍可保证数据的稳定性。

此外，文献数据的可靠性、先进性是由一次文献编辑出版部门控制。文献库的建设不对一次文献做学术评价。从专业角度讲，对确实异常的数据（可能是印刷有误），应予以剔除处理，尽量向用户提供可靠信息。

4、CSJP（英文版）文献库数据质量控制措施

二次文献的加工有多道工序，因此应在各工序加强对数据加工质量的控制。

（1）早期控制

一次文献编辑出版规则和二次文献“规范”，都参照相同的国际、国家文献工作标准。这使二次文献质量控制渗透到一次文献出版成为可能。例如在一次文献中，著者的姓名是否统一用 Guo Zengjian 这种拼音形成。再如，某些核心期刊，文章没有英文的题录与文摘，至使本库无法收录该刊文献，影响了文献数据的完整性。再如个别核心期刊，英文的题录和文摘印刷错误和语法错误较多，这不仅影响刊物自身声誉，也给标引人员带来额外的工作量。努力提高期刊一次文献编辑出版的质量、和规范化标准水平，是保证二次文献数据质量的基础。

（2）比较控制

CSJP（英文版）文献库自 1985 年开始研制，积极参考国内外先进的文献库。国内主要参考清华大学研制的 CUJA（英文版）库，国外主要参考英国“科学文摘”INSPEC 库。1986 年国际联机检索了 INSPEC 库收录的《地震学报》1984 年全年 44 篇文献。将 INSPEC 和 CSJP（英文版）对同一篇原始文献的标引结果进行对比。学习 INSPEC 的标引技巧，对文摘中一些英语语法错误的改正及对长文摘的缩短摘编等。并将一些典型的范例，印发给标引人员参考。

（3）标引控制

文献库质量主要依赖于标引质量。标引就是运用专业知识、文献知识创造二次文献的过程。要求标引人员要熟悉专业、词表、分类法和“规范”，熟悉检索系统和用户的需求，熟练地处理各工序具体的标引任务。所以，标引人员的培训是实施标引控制的主要方法。将分散的、参差不齐的标引状况统一到“规范”上来，是一项重要而艰巨的任务。此外，与标引人员保持经常性的书信联系，反馈工作单中的标引误差，共同探讨标引工作，也是标引控制的有力措施之一。

（4）程序控制

该库工作单的计算机录入，是在编辑下键入的。这有利于提高输入速度和编辑改错，按一定格式录入的数据，用批处理程序变换成 DTR 文件型记录格式。该程序对错误数据类型、数据顺序、文摘过长等有鉴别能力。错误信息存入一误差文件，便于发现问题改正错误。此外，相同数据输入程序、单位名称代换程序等，都着眼于减少数据错误。

（5）人员控制

P. Zunds 曾对标引的一致性做过试验。一篇文献由六人标引，其一致性为 0. 158，两人标引为 0. 543；一篇文献由同一人在不同时期标引一致性是 0. 661（完全一致为 1. 0）。所以，标引人员应相对稳定，不应频繁更换。同时，应尽量减少某个标引工序的多位操作。

（6）总体控制

总体控制是 CSJP（英文版）文献库课题组，对整个文献库标引加工系统进行宏观控制。要

在分散的条件下做到总体控制，最重要的是建立一个具有高度权威和高效管理的控制中心。其任务是：(1) 积极与上级有关部门保持联系，发挥上级机关的职能作用，统筹全局系统的建库工作。(2) 严格执行和维护统一的标引“规范”。包括对“规范”的修订、解释和监督执行。(3) 建立标引工作网。根据情况变化，稳妥调整网中的分工。(4) 有组织、有计划、有教材的进行标引人员培训。日常与标引人员保持工作联系。(5) 掌握建库速度，实施均衡建库。(6) 跟踪国内外情报工作现代化发展动向。研制开发提高数据质量的新技术、新程序，并具体实施使用。

结束语

文献库的建设是一项大规模连续性的信息组织与开发利用工程，同时又是需要人力、物力、财力投入，而见效较慢的工程。它需要各级领导的重视和支持，也需要参与这项工作的全体同志长时期的耐心细致工作。在目前课题经费短缺的情况下，更应珍惜每一条记录的投资，做到建一条成一条。不仅在数量上，而且在质量上将 CSJP（英文版）库建得更好，为我国四化建设和国际情报交流做出应有的贡献。

（本文 1991 年 4 月 10 日收到）

参考文献

〔1〕高崇谦，我国情报检索技术发展与现状，现代图书情报技术，No. 1，P2，1987.
〔2〕黄国俊等，文献数据库质量的计算机控制，现代图书情报技术，No. 1，P5，1990.
〔3〕张希轩，论文献标引，计算机与图书馆，No. 2，P17，1981.
〔4〕邵贻和等，对文献标引若干问题的探讨，计算机与图书馆，No. 4，P28，1984.
〔5〕姜树森，文献工作标准化的概念和发展概况，现代图书情报技术，No. 2，P29，1987.

THE QUALITY CONTROL FOR THE CONSTRUCTION OF THE CHINA - SEISMOLOGY - JOURNAL - PAPER DOCUMENTAL DATABASE (ENGLISH EDITION)

Gao Shuxin

(*Earthquake Research Institute of Lanzhou, SSB, China*)

Abstract

Based on the constructing practice of the China - Seismology - Journal - Paper (CSJP) documental database (English edition) for several years, this paper deals with the significance and complexity of the quality control in the construction of the CSJP documental database (English edition) under the decentralized condition. It presents the construction mode of the documental database, and the quality control mode, control objective and control measures in the process of decentralized indexing for documental data. It is pointed out emphatically that in order to guarantee the quality of the CSJP documental database (English edition), the indexing staffs must adopt firmly the idea of quality first, and it is necessary to establish a control centre with high - degree authoritativeness and high - efficiency administration.

(上接 81 页)

参考文献

〔1〕王泽皋，邢台余震频度增高及以后发生的华北强震，地震学报，Vol. 1，No. 2，1979.

〔2〕王泽皋，关于“震情窗口”问题的实践和展望，地震学报，Vol. 8，No. 3，1986.

〔3〕姜秀娥，华北强地震余震震群应力场“窗口”效应，西北地震学报，Vol. 4，No. 4，1982.

〔4〕敖雪明等，相关地震预报方法的研究，地震预报方法实用化研究文集地震学专辑，学术书刊出版社，1989.

ANOMALY OF FUYUN SEISMIC WINDOW AND PREDICTION BEFORE THE WESTERN HABAHE EARTHQUAKE M7. 3, USSR ON JUNE 14, 1990 AND ITS TWO STRONG AFTERSHOCKS

Wang Guiling, Ao Xueming

(*Seismological Bureau of Xinjiang Uygur Autonomous Region Urumqi, China*)