

摘要

在华夏千年的历史长河中，无数的人与事、地理与天象、工具与科技给中华文明留下了灿烂的篇章，成为中华文明的瑰宝。而中文文献以其独特的形式，在漫长的岁月中记录下发生的一切。过去的那些文章、典籍都承载着历史信息，通过整理与传播中文文献，人们就能对这些信息进行分析，从这些信息中了解中华文化。这篇论文就以中文文献为切入点，涵盖了中文文献的类别、保存方式、搜集方式、研究方式和传播方式等，探讨了整理与传播中文文献的方法。并探索了现代科技对中文文献整理与传播的影响，以北京汉王制造股份有限公司为例，分析了细颗粒度标引的操作流程及发展现状。鼓励大众参与到传承与弘扬中文文献的事业中来，多多关注中文文献的历史价值，积极创新中文文献的传播方式，推动中文文献走出国门，走向世界。文章力求带给人们新的思考，帮助人们推动中文文献更好地发展。

关键词：文献整理；文献传播；细颗粒度标引

Abstract

In the thousand-year history of China, countless people and things, geography and weather, tools and technology to the Chinese civilization left a brilliant chapter, become the treasures of Chinese civilization. The Chinese literature with its unique form, in the long years recorded everything that happened. The past articles and books are loaded with historical information. By collating and disseminating the Chinese literature, people can analyze the information and understand the Chinese culture from the information. This paper takes the Chinese literature as the Cut-in Point, covers the Chinese literature category, the Preservation Way, the collection way, the research way and the dissemination way and so on, has discussed the collation and the dissemination Chinese literature method. Taking Beijing Hanwang Manufacturing Co. , Ltd. as an example, the operation process and development status of fine-grained indexing are analyzed. Encourage the public to participate in the inheritance and promotion of Chinese literature, pay more attention to the historical value of Chinese literature, actively innovate the dissemination of Chinese literature, to promote Chinese literature abroad, to the world. This article tries to bring people new ideas and help them promote the development of Chinese literature.

KeyWords: Documentation; Literature dissemination; Fine particle size indexing

目 录

1 绪 论	1
1.1 研究背景	1
1.2 研究意义	1
1.3 国内外研究现状	1
1.4 研究方法	1
2 中文文献的整理、传播方式	2
2.1 中文文献的整理方式	2
2.1.1 搜集与保存	2
2.1.2 分类与研究	2
2.2 中文文献的传播方式	3
2.2.1 出版	3
2.2.2 大数据检索	4
2.2.3 交流会、研讨会	4
3 北京汉王制造中文文献的整理与传播方式	5
3.1 获取文献资源	5
3.2 中文文献数字化	5
3.3 细颗粒度标引	6
4 中文文献整理与传播的意义	7
4.1 保存文化典籍，记录历史痕迹	7
4.2 方便阅读研究，探寻历史真相	8
4.3 弘扬中华文化，增强文化自信	8
结 论	8
参考文献	10
致 谢	11
附 录	12

1 绪论

1.1 研究背景

中文文献指中国现存的，以中文书写的，并且具有历史意义和研究价值的文章、报刊、典籍等。“文献”一词最早见于《论语·八佾》：“子曰：夏礼吾能言之，杞不足徵也；殷礼吾能言之，宋不足徵也。文献不足故也。”^{Error! Reference source not found.}收集、整理、记录、研究这些中文文献则是中华民族千百年来代代相传的责任。在今天，越来越多的人重视文献，通过文献了解过去，探寻历史的真相。因此这篇文章着眼于中文文献的整理与传播方式，希望中文文献能够以更好的方式呈现在大众面前。

1.2 研究意义

在漫长的历史洪流中，人类会消亡，地形会变化，但文献会将历史铭记，将文明传承。研究中文文献的整理与传播方式，能够更好地帮助人们保存历史记录，留存人类文明，同时也能够帮助人们从文献中获取信息，解析历史问题。研究中文文献的整理与传播方式是帮助人们了解历史，传承文化的途径之一。本文秉持着这种理念，对传统与现代的中文文献整理与传播的方式进行了研究，帮助读者总结其特点，寻找更好地保存和发扬中文文献的方式。

1.3 国内外研究现状

目前越来越多的人投入到中文文献相关的工作中，有许多关于文献整理与保存的文章，人们通常通过文献的整理与保存，研究地方文化的沿革，并分析市场上的出版现状。国内的相关工作者们积极地对中文文献进行研究与出版，国际友人也有关于中文文献整理与传播的著作。国内的中文文献研究视角逐渐走向宏观，而国际上的中文文献研究视角则聚焦于微观，二者相互碰撞，相互交流，将给中文文献带来更好的发展前景。综上所述，目前在中文文献领域研究还有一定的空白，因此本文的研究具有创新性和研究价值。

1.4 研究方法

运用信息研究法，为进行论文研究，笔者在北京汉王制造有限公司进行实习，并负责知识标引的工序。笔者所在的部门，即是将纸质版文献同步为电子版文献，并进行细颗粒度标注的部门，这对论文写作有着实际操作上的帮助。本文将结合北京汉王制造有限公司整理与传播中文文献的做法，探讨中文文献在现代如何继续传承与发扬。

运用文献分析法，围绕中文文献整理和传播的新形态研究这一课题，系统地收集中文文献整理与传播的资料，并借助知网、维普等平台对这些信息进行整合研究。同时大量阅读相关资讯，寻找中文文献整理与传播工作新的切入点，从新的角度探寻整理与传播中文文献的方法。

2 中文文献的整理、传播方式

2.1 中文文献的整理方式

2.1.1 搜集与保存

中文文献的搜集与保存工作是一切工作的开端，做好搜集与保存无疑为后续的研究与传承提供了保障。2022 年 4 月中共中央办公厅、国务院办公厅印发《关于推进新时代古籍工作的意见》，其中提到，提高古籍保护水平，加强古籍存藏基础设施建设，改善保存条件，做好异地、异质灾备保护，确保古籍资源安全。提升古籍修复能力，加强濒危古籍抢救性修复。Error! Reference source not found.

在中文文献的搜集方面，目前大部分中文文献都被中国各地的图书馆、博物馆等专门机构保管，还有少部分在私人藏家的手中，也有小部分散落在全国各地至今仍未被发掘。

大家要积极地搜寻中文文献，首先可以寻求政府的支持，通过政府的影响力广泛地收集中文文献。其次可以通过正规渠道收购，例如出版社、发行商等部门，从他们手中获取原始或者二次加工的中文文献资料。然后通过学术团体、民间机构、旧书商等寻找中文文献，可以有效地提高中文文献收集的效率，同时这类文献能够作为出版文献的补充，对地方文化研究来说能够增强其系统性。最后可以使用网络资源，许多中文文献年代久远，有的已经成为孤本甚至失传，这时网络就能够很好地记录这些中文文献，丰富中文文献的搜集渠道。

在中文文献的保存方面，在历史进程中纸质文献的保存有着极大的困难，随着时间的推移纸质文献会逐渐老化，纸张逐渐变质发生酸化、发黄、变脆，纸质文献无法达到永久保存的目的，所以在对其进行原生性保护的同时，也逐渐开始了次生性保护。有资料表明，民国普通报纸的保存寿命为 50 年左右，图书的保存寿命为 100 年左右。目前，酸化已成为危害近现代文献的主要杀手，全国所有近现代纸质文献都不同程度地受到酸侵蚀。

随着现代科技的进步，中文文献保存的手段也有了创新，包括对传统除虫、温度控制等技术的创新，以及古籍修复、古籍再造等技术。通过应用这些技术，中文文献的保存也更变得加便捷化、多元化、网络化。书籍是人类文明的火种，保护中文文献是每个人义不容辞的责任，中华儿女要肩负起历史的重担，重视中文文献，切实保护好中文文献，为后世留下具有意义的记录。

2.1.2 分类与研究

通过对中文文献的分类整理，能够对收集来的资料进行分门别类的整合，梳理文献之间的联系。对中文文献进行研究可谓是中文文献存在意义中最重要的一点，通过研究中文文献，能够了解不同时代的社会

风貌，了解不同地域的风土人情，上晓天文，下知地理，帮助大众拓宽视野，拓展知识面。

在中文文献的分类方面，文献分类是指根据文献内容和形式的异同，按照一定的体系有系统地组织和区分文献。中文文献的分类为后续中文文献的研究打下了基础，通过分类大家才能更方便、更快捷地找到自己所需要的资料。分类的方式有许多种，包括按时代划分，按地域划分，按内容划分，按作者划分等等。举例来说，如果按照时代划分，《明史》《三国演义》《水浒传》《西游记》都属于明朝的文献，《东方杂志》《清华周刊》《万有文库》等则属于民国时期的文献。

如果按照中文文献的载体来分，包括纸张文献、微缩文献、电子文献、音像文献四类。纸张文献是以纸张为载体，用书写或印刷等方式记录知识的文献。这种以传统手段保存的文献容易损坏，不利于长期保管。微缩文献是用缩微摄影技术把印刷品或手稿按比例缩小而产生的文献形式，这种文献便于携带但需要专业的读取设备。电子文献是指以数字代码方式将图、文、声、像等信息存储在磁光电介质上，可以使用计算机等设备读取的文献类型，这种文献是现如今应用非常广泛的一类文献。音像文献是以磁性材料、光学材料等为记录载体，利用专门的机械装置记录与显示声音和图像的文献。人们通过观看图像、视频，或聆听声音来读取信息，这种文献生动直接，能使读者更直观的获取信息。

在中文文献的研究方面，对于带有明显地域背景类的作品，研究者可以从中发掘地方的风土人情及地理沿革，还可以研究地形地貌的变化以及地址方面的信息。例如陈忠实的《白鹿原》，高满堂的《闯关东》，贾平凹的《秦腔》等。

对于着重描写历史变迁，时代变化的作品，研究者可以从中探寻历史的演变特点，发掘特定时期的政治、经济、文化发展状况，了解当时的时代背景与制度政策等。例如梁斌的《红旗谱》，曲波的《林海雪原》，老舍的《四世同堂》等。

对于家族之间流传的家谱、族规等作品，研究者可以从中了解一个家族的人员增减，繁荣与衰落，以及聚居地的迁移，还可以得知当时时代的风俗礼仪以及官职制度等。例如《湖南会同吴氏族谱》《重庆云阳黄氏族谱》《绥远县志》等。

2.2 中文文献的传播方式

2.2.1 出版

在中文文献的传播方式中，对文献进行加工出版是较为传统和普遍的一种方式。出版又称发表，是指将作品通过任意方式公之于众的一种行为。作品获得国际标准书号并经过一定资质的出版机构印刷成书籍的称为出版物，出版物内容以数码形式呈现的叫做电子出版物或电子书。拥有大量复制或同等规模的传播力的公开发表都统称为出版。

这其中包括对原著文献进行校对、审核、编辑、排版等工作，保证出版的中文文献的准确性。有时也会邀请画师，在文献内容中加入插图等，方便读者理解，同时增加阅读兴趣。例如《金陵全书》，由南京出版社出版，包括县志、墓葬、政府管理自信等诸多内容。

一方面，中文文献的整理出版具有重要的学术和现实意义，能够突破旧有的自上而下的研究视角，拓展历史事实的研究。学术界与出版界通力合作，在前人的基础上系统辑录、整理资料，供将来的研究参考和使用。另一方面，中文文献的出版有助于传承历史与文化，加强人们对中文文献的保护意识，同时为多种学科的研究提供依据。

中国各地的出版社、发行商为中文文献出版事业贡献很大，例如大连出版社，是国家推动出版业改革的代表之一，是由国家出版总署批准，在部分副省级城市和计划单列城市中成立的 16 家城市出版社之一。大连出版社已累计出版图书 3000 余种，涉及 80 多个门类，切实推动了中文文献面向大众工作的发展。将中文文献重新整理，以全新的面貌呈现在大众面前，能够使民众更方便地阅读资料，也能够使中华文明更广泛地传播，更长久地延续下去。许多中文文献出版物已远销海外，被翻译成多种语言，使世界更加了解中国，了解中华历史的传承，了解中华文化的发展进程。例如《浙江文丛》《江苏文库》等大型出版工程项目。

2.2.2 大数据检索

在科技逐渐发展的今天，中文文献的传播方式也不仅仅局限在纸张上。文献检索是以科学的方法，利用检索工具和检索系统，从有序的文献集合中检索出所需的信息的一种方法。中文文献的传播也依赖于网络空间，人们将中文文献上传至网络，以数据的方式保存文字信息，同时各部门建立网站、APP 等，通过检索的方式查找所需的信息。

例如“知网”，即中国知网知识发现网络平台，它面向海内外读者提供中国学术文献、外文文献、学位论文、报纸、会议、年鉴、工具书等各类资源统一检索、统一导航、在线阅读和下载服务。还有许多地方政府文教部门、图书馆等部门机构也会建立自己的网站，通过互联网就能够查询到文献的信息，包括页数、出版单位、发行单位、语种等等。例如在沈阳图书馆的官网上就能查询到《沈阳地方志》的相关信息，并且还能进行预约、查找馆藏等操作。

通过大数据检索的方式，不仅能够促进信息资源的迅速开发和利用，而且能够帮助科研人员继承和借鉴前人的成果，避免重复研究，少走弯路，节省查找文献的时间，进而加速研究进程。在网络连接世界的今天，外国友人也能够通过互联网阅读中文文献，更直观详细地了解中华文化。互联网使中文文献走出国门，融汇进世界文明的大浪潮中。

2.2.3 交流会、研讨会

开展中文文献的传播工作，是一个人与人之间交流并交换信息的过程，可以促进社会团体的发展。我们可以定期举办一些交流会、座谈会与研讨会，确立一个主题，邀请对主题有研究经验的相关从业者，一起来探讨问题，交流想法。这样中文文献的信息就在人与人之间传递了，进而达到了传播的效果。这种以口头或书面形式进行的中文文献交流与传播，具有直观性和针对性，人们以线下或线上面对面的形式进行信息交换，能够直接地了解到其他人的想法，得到不同视角的观点。

例如，2022年在四川大学哲学系举办的第四届全国画学文献学术研讨会，这次研讨会主要探讨了中国画学文献与美术史的研究，画学文献与图像等。2021年在曲阜师大举行了《论语文献集成》发布会暨“论语文献研讨会”，会上庆祝了《论语文献集成》的出版，并讨论了《论语》文献的选择标准、文献整理分类、文献研究等问题。

通过问卷调查的结果能够发现，目前人们对中文文献的整理方式与传播方式有着一定的了解，但了解程度还不够深。这也反映出中文文献的宣传力度不够，人们没有渠道，或者没有兴趣了解中文文献的问题和情况。因此应该加大宣传力度，多多开发更加便捷的阅读与查找中文文献的方式，为中文文献的传播与推广做出贡献。同时也侧面反映了整理与传播中文文献具有特殊的意义，能够使人们更加了解中华文化。

3 北京汉王制造中文文献的整理与传播方式

3.1 获取文献资源

笔者的实习单位为北京汉王制造股份有限公司的子公司——汉王数字股份有限公司，在实习过程中笔者接触到了该公司对中文文献进行整理与传播的方法。该公司的主营业务是将纸质版文献转化为电子版，并进行信息整合。

北京汉王制造股份有限公司与全国的图书馆、博物馆、出版社、学术团体、收藏家、律所等机构和个人接洽，将包括图书、期刊、杂志、报纸、表单、公文等在内的纸质版中文文献转化为电子版中文文献。随后，制作好的数据可以应用在学术机构的网站和 APP 上，供人们检索阅读和保存下载。通过与客户的对接，公司就取得了第一手资料。部分中文文献将通过邮寄的形式运送到公司，再由员工进行下一步整理；部分保密的文献不方便邮寄到其他地方，将由公司外派员工，到客户所指定的地点进行文献整理，以确保文献资源不外泄。

3.2 中文文献数字化

不论是邮寄到公司的中文文献，还是在客户指定地点进行操作的中文文献，拿到中文文献后进行的第一个工序就是对其进行扫描。使用扫描仪和 PDE 软件进行配合，将纸质版文献转化为电子版文献，包括 PDF、Html、Txt、Doc、Xls 等多种形式的文件。并且要注意调整文件的阈值、对比度、亮度等，这关系到后期文字识别的工作。还要注意不要出现缺印、漏印、模糊的情况。在扫描中文文献的同时也要注意对中文文献的保护，不要对其进行破坏，尽量保持中文文献的原始状态，对于有些中文文献中携带的便签、签条等也要注意保存，不要丢失。这一步使得实物文献转移到了计算机上，成为了数据信息文献。

然后将扫描得到的电子版中文文献进行图像处理，即用计算机对图像进行分析，以达到所需结果的技术。图像处理技术的一般包括图像压缩，增强和复原，匹配、描述和识别 3 个部分。对电子版中文文献进行去黑边、去杂点和校对的处理，并注意检查文件的清晰度和文件与页码的对应关系，为后续的工序打好基础。

随后进行版式划分。版式即版面格式，具体指的是开本、版心和周围空白的尺寸，正文的字体、字号、排版形，字数、排列地位，还有目录和标题、注释、表格、图名、图注、标点符号、书眉、页码以及版面装饰等项的排法。将前面得到的文件用框选的方式划分出文字与图片，调整不同分框之间的顺序，这样文字才能按正确的顺序排列。同时进行生成目录的操作，将中文文献按“篇”划分为不同的层级，按照这个层级生成目录，目录中包括了栏目、子栏目、篇目的层级，为后续对中文文献进行细颗粒化标注做好了准备。

接下来使用由汉王制造股份有限公司研发的 OCR 技术对文字进行识别。OCR 即 Optical Character Recognition

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/108001077011006054>