

# 把中华文化传遍全世界

## 《四库全书》文津阁本数字化标准版完成

最近，在习主席“以古为鉴”，“开创未来”的号召下，在严重的疫情中，我们完成了历史最大的数字文化工程，拥有 8 亿字以上的文津阁《四库全书》的全文数字化。它浩如烟海、博大精深、包罗万象，被誉为与中国古代的万里长城，京杭大运河比美的文化工程。只有把它全文数字化，才能把中华文化的大集成，传遍全世界。

在清乾隆时期，中国经一百多年的励精图治，清朝呈现出三十七年的（1772 年）经济持续发展，国库充盈，无内忧外患的盛世景象。雄才大略的乾隆皇帝据个人喜好，在学术界的呼声下，中国传统的盛世修书在这时候又展现出巨大的魅力，催生出继顺治、康熙、雍正之后再次修纂大型书籍的态势。盛世的宏阔是需要鸿篇巨制来充实的，而清政府也明白需要完成时代留给自己的使命。

乾隆在 1772 年 11 月，他认可学政朱筠提出明永乐年修的《永乐大典》的辑佚问题，便诏令将所辑佚书与“各省所采及武英殿所有官刻谱书”汇编在一起，引导出了编纂《四库全书》的前无古人的巨大文化工程。

《四库全书》的编纂过程中，第一步是征集全国古籍经典。在乾隆帝亲自下诏下，从乾隆三十七年（1772 年）开始，至乾隆四十三年（1778 年）结束，历时七年之久，进展顺利，共征集各古籍经典 12237 种，其中仅江苏进书就 4808 种，浙江 4600 种。在乾隆亲自主持下，由纪昀任总编纂，集中 360 多位高官学者编撰，3800 多从全国竞选出来的高手抄写，耗时十三年编成。分经、史、子、集四部，故名“四库”。共收录 3462 种图书，共计 79338 卷（相当于《永乐大典》的 3.5 倍）。36000 余册，约八亿字，单字则有楷书 8 万字以

上。

为保证质量，又组织多名纂修官审校，计有分校、复校、总校有三校审后，最后呈乾隆御览审校。

《四库全书》共有 79000 多卷，36000 多种多册，分“经”、“史”、“子”、“集”四部。其中“经部”分为“易、书、诗、礼、春秋、孝经、五经总义、四书、乐、小学。”等 10 类。以易类为主。仅小学类中，就有 80 多部世上罕见的古字典。“史部”分为“正史、编年、纪事本末、别史、杂史、诏令奏议、传记、史抄、载记、时令、地理、职官、政书、目录、史评”等 15 类；以正史为主。“子部”分为“儒家、兵家、法家、农家、臣家、天文算法、术数、艺术、谱录、杂家、类书、小说家、释家、道家”等 14 类；“集部”分为“楚辞、别集、总集、诗文评、词典”等 5 类。总共 44 类。它包括所有经、史、子、集，四书五经，诸子百家。二十四史，各种古代医书，各种古代字典，古琴谱，农业、机械、战争、科技无所不包。甚至还有似“清明上河图”的 119 双幅页，《八旬万寿盛典》图，行军布障图、狩猎图……无所不包。

《四库全书》浩如烟海，包罗万象，它是“中华传统文化之总汇，古代典籍之渊鉴”。它最能代表中华文化博大精深的载体的伟大的文化工程；正因为它是中华民族乃至全人类的数千年的思考、智慧的精华，也因为它是乾隆皇帝亲自统领完成的无可比拟的国宝，所以自它诞生以来，一直被精心、严密、保管锁在深宫大院，只能远看，不能近翻。就是历史学家的研究者，也很少看到它的一部分。为使她能扩大服务于读者，二百多年来，多少历史名人、专家、学者、达官贵人，都处心积虑前后 100 多年间，五次希企以影印出版，但都因技术、资

金、战争……都未能实现。

在 1986 年，即《四库全书》完成的 208 年。由台湾商务印书馆完成了 1500 册一套的影印《四库全书》文渊阁本 500 套。它的版面虽压缩后上下各一半版面，就达 1500 册，16 开本，每册约一寸厚，1500 册摆起来，也有 50 米长。分期完成，后各由全世界各大图书馆、博物馆所购，而分散到全世界。然而仍都做为镇馆之宝，严加看管，仍未起到服务于大众，光大中华文化的作用。

又在 2003 年，上海古籍出版社又照台湾版影印 1500 册，32 开本 500 套，发行之后，还是太浩大。又到 2006 年，对《四库全书》早有百年情结的商务印书馆，又用扫描技术，印刷 1500 册的《文津阁》四库全书。

同样，她们浩大、辉煌，但同样难以阅读、检索等研究。在这期间的 1997 年，曾出版过一个《文渊阁》四库的全文电子版，但由于当时的计算机中文信息技术水平所限制，仍很不标准，不成功。

### 世界华人发明博览会

1998 年 9 月，在香港新建成的大型会展中心，召开了由中国发明协会牵头香港地区、台湾地区发明协会，美国华人发明协会，共同主办了聚集世界华人的优秀发明 2000 多项的“世界华人发明博览会”召开。中国发明协会在主席聂力中将带领下，携近 10 年中国的优秀发明 600 多项参会。此会震动香港和各地华人，参观者在香港这商业特区，表现的空前踊跃异常。博览会由来自世界各地的技术发明专家组成的评委会共 15 人（中国大陆 2 人），每天穿梭于各展台进行观察、了解、评估。最后在隆重的发奖大会上，由香港特首董建华先生亲自颁奖。获得大奖第一名的被公认的是北京中易公司参展的解决了

一切汉字信息处理的《超大型全汉字信息处理集成》项目。它是著名的《英华大词典》主编郑易里对汉字钻研半个世纪，后又由我们三代人，半个世纪完成可对十万以上汉字中文的字库设计、《郑码》，它可对任何汉字检索、排版、系统扩充的全部的中文信息进行有效处理。突破了利用计算机处理一切汉字信息的历史难题。用事实打破了世界上认为汉字进不了计算机，汉字不能信息化的错误认识。历史上曾有无数的包括鲁迅这样的专家，都纷纷要求中华民族废除汉字，而改文字拉丁化的形势下，使汉字跟上了世界信息化前进的步伐。用事实有力的粉碎了“汉字必须拉丁化”的谬论。同时，它反而证明汉字是世界最古老，最优秀文字，而又是最适合计算机处理的文字。它振奋了所有参展华人发明家！在宣布此项目获得第一大奖时全场起立，热烈鼓掌祝贺。中国代表团团长聂力中将也难以掩饰兴奋的心情，热烈祝贺中易公司的代表，为中国大陆发明者增光！为世界华人增光！

随后，这个项目又获得北京科技进步一等奖，国家科技进步二等奖等国内外重大奖项。

这个项目解决了对任何汉字信息支持文书打印、印刷处理。随后，我们为汉字信息传遍全世界，又向符合有全部国际标准 7 万多汉字的大型国际网络平台进一步研发。它需要解决由二字节系统向四字节系统的扩充；支持四字节的搜索引擎；支持四字节全文检索数据库和大型管理发布系统。为我国的中文信息处理大量古籍数字化打下了雄厚基础！

### 《国家 973 信息技术与高性能软件规划项目》

2001 年 11 月，我们的《超大型中文网络国际应用平台》项目被国家科技部选列为“国家重点基础研究发展规划项目的《国家 973 信

息技术与高性能软件规划项目》，它要支持服务器端、客户端、系统扩充，还要跨语境，跨平台，跨系统，还要有能迅速对世界一切文字，对 9 万不同的有国际编码的标准汉字不但要有能快速搜索输入的《郑码》系统，还要有能快速检索的功能强大的搜索引擎。这个搜索引擎不同于当前“百度”，“新浪”等网站建立在搜索几千常用字的，由 2 字节系统设计成的搜索引擎。因它要用于古籍数字化产品的快速准确的检索，分类的全部功能，所以它必须设计成支持 4 字节被大大扩充的操作系统。还要针对古籍特点建立规则的索引，智能推荐，简繁体字的关联检索等等独创性的设计。这才能在全世界近 200 多种国家和民族语言文字，在国际互联网上畅通无阻。鉴于它对光大中华文化的重要性，此项目由国家科技部 IT 领域的首席科学家顾钧亲任该课题组长。由于我们在此项目前，早已有深厚基础，所以在不到两年时间便顺利完成，随后通过国家主管部门测试合格和美国微软的认证。成为《国家 973 大型信息基础项目》中，完成并应用的完成最快，最顺利的项目。由于它的完成，使所有国际标准编码的近七万汉字焕发青春，恰当其时地创新工程使全部汉字信息处理，领先于计算机文字信息处理的高端。

2003 年末，国家图书馆计划，要开展国家数字图书馆的首个大型 250 万页的古代地方志全文数字化工程。他们在我们和另一大型中文信息公司之间进行选择，经过前后一年多，认真、严格的考察了，最后选中我们，国图的评价很简单是我们“事业心强，技术全面”。在随后工程合同中，国家图书馆强调“全部数字化工程，要建立在《国家 973 信息技术项目》”之上。致使完成后，可以跨系统，跨平台，能在全世界浏览和检索。为全世界人类服务。

古代“地方志”，由于是由各地乡绅出资，自行编写，由于各地水平相差极大，各种生僻字，异体字，自造字，讹字无处不在，文字书写水平，有的木刻版的拓印水平也参差极大，它是最困难的信息处理的古籍资料。由于字迹太差，由于 OCR（图形识别软件）识别率太低，只能由操作员用输入一切汉字的《郑码》来敲键输入。

两年中，我们没有辜负国家图书馆专家的信任。突破了一个又一个的前所未有困难。当这一复杂巨大工程的数字化工程，分批交国家图书馆组织的五个部门的专家分批验收中，每一次都检验合格，不断获得各部门专家的高度好评。

没有这《国家 973 信息技术与高性能软件规划项目》的全面、创新、升级功能匹配的完成，就没有《四库全书》全文数字化工程的实现。

## 奋力《四库全书》全文数字化

我们几十年如一日的向汉字大字符集的突破性研发，企求的目的就是将中华文化几千年的民族智慧，古籍经典逐一实现数字化；就是致力不断实现“以古为鉴”，将几千年中华民族的智慧通过国际互联网传遍全世界的梦想。让全世界认识中国的古老文化和中华民族的智慧！

所以我们早就瞄准了中华文化的“皇冠”——《四库全书》的全文数字化。

我们是个小而强的公司，只有很少的资金，只能是靠勤奋工作日积月累一点点的，以“滴水穿石”的韧性，持久力，经过了 14 年，完成了 90%的《四库全书》文渊阁本的全文数字化。而剩的 10%，则是最难，最复杂，往往是当今中文信息技术很难达到的古代 80 多部

的古代大字典。

对《四库全书》还情有独钟的，就是有 130 多年历史的商务印书馆，他们从创始人张元济发明照相排版技术后，就耿耿于怀的希冀将《四库全书》影印出版。让更多的读者看到四库全书！在这一百多年间，也有无数名人如章士钊、张学良、叶恭绰等多位和政府单位，都几次发起影印《四库全书》。但是，都因“需费浩钜”，“工程浩大”，“战乱不稳”等等而半途而废。直到 1985 后，才有由台湾商务印书馆的影印《四库全书》文渊阁本陆续出版。随后，上海古籍出版社影印文渊阁版出版。北京商务印书馆影印《四库全书》文津阁本的出版。这些影印版，仍由于部头都在 1500 册以上，各方仍以珍贵而深藏不露，使专家仍无从着手的查找，浏览和研究，仍起不到这中华民族文化“以史为鉴”，让世界认识中国的作用。

当然，还是商务印书馆对《四库全书》的执著和忠诚，在出版了“文津阁”《四库全书》影印版后，认定只有出版《文津阁》现代全文数字版，才能充分焕发《四库全书》的青春，才能将其从深宫大内走向国际互联网，走向全世界！申请了国家重大文化工程项目的资金，但却一直搁置多年，为找不到实施这巨大、艰难工程的单位而焦虑。当他们得知中易公司早已开始艰难地将“文渊阁四库”进行全文数字化即将完成，以后，随即开始了对中易公司的工程考察。直到招标、签约。然而，签约工程正式开始后不久，严酷的新冠疫情已席卷而来。

### 文津阁本、文渊阁本《四库全书》的不同

文津阁本具有独特的历史文献价值。从七部《四库全书》成书时间上看，文津阁是北方四阁中最后完成的一部，距第一部文渊阁本，成书后，有三年之久。

它是档案明确记载曾经乾隆亲校，并由纪昀亲自三校的抄本，几乎每册均有校核纪录，其编校质量优于包括文渊阁本在内的其他诸本。

由于战乱，文渊阁本几次搬迁，唯独文津阁本保存在承德“避暑山庄”。后自1914年迁至北京后藏于京师图书馆，80年代移至北京图书馆新馆，长时期未经移动。因此文津阁本保存最为完整，其他三阁本均分别据文津阁本加以抄补。如1917年，贮存于清内务府的文渊阁本，经检查，有9种书缺佚，共缺23卷，反据文津阁本补抄。1934年，藏于沈阳的文溯阁本，发现缺卷，也据文津阁本补抄了《挥麈录》等三书。杭州文澜阁，则据文津阁本补了更多。可见文津阁本在历史上已起了相当大的文献补辑作用。

20世纪90年代初，国家图书馆研究员杨讷先生主持了部分文渊阁《四库全书》影印本与文津阁《四库全书》原书核对录异的工作，出版《文渊阁四库全书补遗（集部）》，共15册，二百余万字。（仅差异就写了200万字，可见差异量之大，差异处之多），其中历代诗文就补四千余篇。全部辑自文津阁本只集部书，为同书之文渊阁本所未见。（仅据该书推测4000篇\*100字=400000字，集部补入录校工作的工作量不低于40万字），据统计，文渊阁本集部共收书1273种，其中与文津阁本有差异的有788种，而宋代诗文集，文渊阁本失收、可据文津阁补入的，有1160条，涉及118种书。两者差异达20%左右。这些差异包括篇卷差异、文字差异、序、跋、附录的差异、提要的差异等。尤其是，文津阁本保存了《永乐大典》的内容，所辑佚的内容远较文渊阁本丰富。通过核对差异还发现，文津阁《四库全书》所保存下来的一些文献资料竟是海内孤篇，尤为宝贵。

文津阁本，由于保存在承德“避暑山庄”，据说乾隆有些年。会

在那 8—9 个月的时间。常翻阅这文津阁《四库全书》。所以文津阁本还有乾隆阅读时书写的近 3 千条御批，这些御批，字迹恭整，涉及阅后感、批示、感慨、纠正等等内容，很广泛，这也是其他版本所没有的。

尽管历代很多学者努力阅读部分《四库全书》，但是由于篇幅太大，借阅太不便，研究的还只是部分。至今，只有采用计算机对全部内容进行校勘比对，才会使以前不可能完成的工作具有可行性。但是需要至少两个先决条件：1，同时拥有文津阁本和文渊阁本的全部电子全文数据；2，设计具备计算机校勘比对系统。我们在过去十年间，持续对文渊阁本《四库全书》完成全文数字化。现在已具备专家们对这两个版本进行全部校勘的条件了。当然这两个版本的不同之处还有许多，在此就不一一深叙了。文津阁本《四库全书》全文数字版的标准化、全文化、网络化的完成，实现了供全世界浏览和检索。

《四库全书》不仅是中国，也是世界上唯一的一部空前的、最早、最大的一部百科全书巨著。它集中国古代典籍，中国文化之大成，对保存中国古代文献作出了巨大的贡献。而全文数字化后的数字网络版的完成，才能使它传遍全世界，从而它不但能增强中国人的骨气和底气，让世界更好认识中国，了解中国，需要深入理解中华文明，从历史和现实，理论和实践相结合的角度，深入阐释如何更好坚持中国道路，宏扬中国精神，凝聚中国力量。

联系方法：

电话：13801228838

地址：北京朝阳区北四环东路 108 号

乙 5 号楼 1004

E-mail: cstc@china-e.com.cn

北京中易中标信息技术有限公司

蓝德康

2023 年 月