

龍語瀚堂典籍數據庫



www.dragoninfo.cn

終結古典文化研究的影印歷史！

打破古文字無法編輯的現實堅冰！

開創四字節典籍研究的數字時代！



社會科學文獻出版社



北京龍威特信息技術有限公司



北京時代瀚堂科技有限責任公司 聯合制作

概述

■ 古籍整理与研究的现状

随着人们对传统文化的重视和计算机技术的发展,利用信息技术对古籍数据库建设的工作得到高度重视和快速发展。众所周知,传统的二字节编码技术只能处理2万多汉字,而汉字种类约有7万,这还没有包括甲骨文、金文、简帛等出土文献研究中使用的无法求定的汉字。因此,以往的计算机技术在使部分古籍实现了数字化的同时却依然对大量生僻字、古字束手无策。研究者在遇到生僻字的时候,往往利用造字、符号表示、制作卡片、利用语言对字的组成的描述等手段来表述,从而给学术间的交流和传播带来极大不便。

■ 古籍数据库应用的现状

机构自建和购买是目前古籍类数据库资源建设的两个主要渠道。在肯定现有大量数据库对古籍数字化贡献和典范作用的同时,不可否认的是技术瓶颈——对生僻字的处理没有达到完全数字化,使其在内容方面有着不可避免的缺陷——对生僻汉字出现频率较高的字书类、小学类、音韵学、训诂类、金文类等古籍的处理均被长期搁置。

■ 舍弃生僻字而造成中文知识库的不完整

据权威统计,在一般古籍中约有1%的汉字属于生僻字,而在诸如《说文解字》等小学类古籍中,出现频率甚至高达3%以上!汉字经过数千年的发展和演化,目前流传下来的种类总量约有7万左右,大量的汉字由于使用人群和领域的狭窄,被归结为“死字”。但是这些生僻汉字大量的存在于古代典籍中,与常用汉字一起构成“汉字”这一中华文明的载体,它们对于历史和文明的考证有着不可或缺的作用,其本身也是文化遗产的重要组成部分。

因此,舍弃了生僻字而建构的这些中文知识库都不是真正完整的!



■《龍語瀚堂典籍數據庫》標志着漢學研究進入四字节時代

《龍語瀚堂典籍數據庫》採用基於Unicode四字节編碼和自然語言全文檢索的典籍文獻數字化構建技術，解決了四字节實際應用中的一系列相關問題，从而使四字节技術真正得到開發和應用，不但結束了古籍研究手工抄寫、圖片替代、生硬造字的歷史，也標志着漢字處理開始進入四字节字域。

由於該系統使用了國際通用的Unicode文字編碼，使它不但能夠支持二字节編碼的常用漢字，而且徹底突破古籍整理和研究中生僻漢字數字化的瓶頸，開創性地實現了對所有漢字通用的典籍文獻數字化。

■ 古籍研究的必由之路

—借助現代計算機技術，建立在數據庫層級

以往的研究過程中，無論是學生還是老師，總是處於逐本翻書查找資料、手工加書籤、做卡片的过程，更何況面臨大量的造字工作。而當關於該學科所有的相关資料和典籍集成庫的時候，使用者不但可以自由的瀏覽、檢索、下載和編輯，同時數據庫本身的統計功能更可以使以往繁複的查詢工作在毫秒間完成，大大減少研究者的工作量，例如可以瞬時得到關於《康熙字典》中對於《五音集韻》的引用次數是948次。除此之外更為可貴的是，建立在數據庫層級的研究往往會延伸出更加獨特的角度和發掘出更多的研究領域。

使用

■ 輸入方法：

龍語瀚堂典籍數據庫頁面提供專門針對生僻漢字的輸入方法，解決研究人員最基本的录入問題。新概念五筆輸入法是目前唯一支持unicode超大字符集的五筆輸入軟件。該輸入法在www.dragoninfo.cn上提供下載。

■ 檢索方法

無論用戶從首頁選定數據庫還是直接開始檢索，系統經過分步檢索，從“總庫—指定數據庫—指定書目—內容顯示”三個環節得出最終全部資源內容。而在任何環節中，用戶都可以重新選定某個數據庫，或者直接返回首頁面對總庫進行關鍵字的檢索。



■ 軟件要求：

操作系統：Microsoft Windows 2000或以上版本

軟件：Microsoft Office XP 簡體中文版或以上版本

瀏覽器：Internet Explorer 6.0



特 点

■ 具有在通用浏览器上自由编辑和浏览的功能

建立在Unicode超大字符集之上的《龙语瀚堂典籍数据库》采用四字节编码技术使计算机可处理的文字种类数达到7万之多，即彻底解决这些四字节汉字在计算机平台上的录入、显示、编辑、检索、查询和管理等问题。全库无造字，所有文字可以在MS-office上进行复制、粘贴和通过互联网进行发布，从而结束了古籍研究手工抄写、图片替代、生硬造字的时代，也为研究者、古文化爱好者查阅、认知古籍原貌提供便捷途径，使得所有古籍类图书信息化成为现实，对于中华文化的保存和发展具有极为重要的意义。

■ 独有的四字节自然语言全文检索技术

相对于目前基于二字节字符系统的（即UCS-2）的互联网全文检索，《龙语瀚堂典籍数据库》www.dragoninfo.cn是全球第一个支持四字节字符（即UCS-4）、基于互联网的自然语言海量文本全文超高速检索系统。该引擎可以从海量的含有UCS-4字符的文本中在毫秒量级内进行信息抽取。采用B/S模式，在微软平台上管理的字符总数达到7万个。

■ 内容自动累加成库

该功能无论是对于历史、博物馆界、图书馆界、还是对于各种的研究机构或个体的研究者，都可以通过它来建立一个与自己研究领域或方向相关的专业数据库。无论是

图书馆还是出版社均可以利用此项功能实现其从逐本图书资源的提供者转变成海量且系统性的数据库资源供应者。

■ 众多的专利技术

由于在四个字节字符方面拥有的核心技术专利池，和对含有四字节汉字内容的版权群，形成了对以UCS-2为技术核心的目前大量古籍数据库及其它检索网站的战略优势，为其主导甚至全球范围的网络检索领域的发展、建构领先的超级知识库提供了坚实的技术保证。龙语瀚堂典籍数据库已经和正在申请的核心专利包括：

生僻汉字全信息输入法

书刊文献数据库及其光盘制作服务系统

对含四字节文本进行全文检索的方法

四字节网络检索的方法

含四字节字符网页抓取和管理方法

含四字节字符文本编辑器

四字节桌面检索系统





內 容

建立在先进技术基础上的龙语瀚堂典籍数据库是一个专门基于四字节汉字处理的古籍数据库。其本身自带了汉学研究所必须用到的《说文解字》、《康熙字典》等字书类数据库。所有内容都是因为在此之前由于四字节处理技术瓶颈未被突破而无法在计算机平台上实现数字化的内容。除此之外还有大量的金石拓片，并同时尝试蒙文、藏文、梵文等其他民族语言和文化的整理。

数据库名称	内容涵盖	说 明	功 能
字书类数据库	《说文解字》(上、下册)、《康熙字典》、《尔雅》等	提供多种检索方式，繁简自动转换，与总书对照，全库无遗字。所有字书类数据经过精粗校对，达到万分之零点八的出错率。并配套出版纸本图书	全文检索 浏览功能 版本对照 导出编辑 原文打印
中国古文字库	以单字的面文并线的方式呈现中国古文字内容，包括每个汉字从甲骨文到楷书其演变过程的所有信息	2005年内推出《说文解字》、《行书》、《古文四声韵》中的古文。其他内容陆续上网	
殷周金文库	《殷周金文集成释文》、《金文集成》未收者(81年以前)，伪作伪刻漏未等	释文并释全文，并陆续提供器物照片、线图、铭文拓片、铭文摹本、器物图典数据等内容	
中国古钱、古印库	收录战国、汉代、明清以来的印章、古币	提供拓片、行释、出处等内容	
考古、文字学书目库	收录所有考古学、文字学的相关书目与书目资料	包括书、刊名、作者、出版社等相关信息	
简牍库	基于互联网的所有关于简牍研究的文集集成	2005年内推出	
台湾国学报告数据库	包括考古、历史、文字学等人文学科的台湾政府研究报告与学术论文	2005年内推出	
中国音韵库	包括《集韵》、《广韵》、《五音集韵》等传版韵书，以及音韵学领域学术论文和罕见古韵书	2005年内推出	
小学类数据库	《说文解字》、《说文解字注》、《说文解字通定》、《说文解字通考》、《说文解字通考》等，是研究《说文》系列的重要数据库	目前提供段玉裁《说文解字注》全文，以及《说文解字通定》、《说文解字通考》展示版本。其他图书陆续提供上网	
金文文献库	包括中国近现代金文研究的历史文献和学术专著	2006年中推出	
古籍核心期刊库	包括目前古籍研究的核心学术期刊及学报	2006年中推出	
甲骨文库	包括中国近现代甲骨研究的历史文献和学术专著	2006年中推出	
中国古籍库	包括中国历史典籍、《十三经注疏》、《二十五史》等常见古籍	2005年至2006年内陆续推出	



龍語瀚堂典籍數據庫

目前唯一可在网络平台上支持超大字彙集，可进行自然语言全文检索的，突破编辑功能的典籍类数据库。

2005年6月，龙语瀚堂典籍数据库中的《说文解字》图书及数据库网络版在中国社会科学院隆重发布。国务院古籍办、新闻出版署、科技部、社科院历史所、社科院文献中心、出版科学研究所以及图书馆界的与会专家学者高度肯定了此项工作对于古籍数字化处理的变革意义。《光明日报》、《科技日报》、《出版参考》、《图书商报》、《经济导报》、《新京报》、《京华时报》等十多家媒体也对此次发布会作了详细报道。随后，人民网、中国新闻网、新浪网等多家网络媒体也对此进行了大量的转载报道。



Dragoninfo Technology Co., Ltd.
北京龙戴特信息技术有限公司

地址：北京市海淀区世纪城蓝靛厂东路2号楼金源时代商务中心A座10E 100089
10E, A JinYuan Times Business Center, Building No. 2 Landianchang
East Road, Shijicheng, Haidian District, Beijing, China. Zip: 100089
电话 (TEL): 010-88681500 51653993-836
传真 (FAX): 8610-88861600
WWW.dragoninfo.cn
客服信箱 (E-MAIL): cs@dragoninfo.net.cn