

甲骨及甲骨拓片影像数据库的设计和实验

贾双喜

甲骨是在河南安阳西北的小屯村发现的，至今已有 100 余年。到目前为止已出土甲骨 15 万片左右，分藏于国内各大博物馆和图书馆，也有许多流散到世界各地，成为中国的海外遗珍。甲骨文不是中国最早的文字，它只是商代晚期的文字。世界文字学家认为，一种成熟的文字一般要经过 2000 年的发展，而 3000 年前的甲骨文已经出现了一批形声字，从这一点看足以证明它也是比较成熟的文字。在已发现的近 5000 个甲骨文单字中，能够释读的约 1500 字，余下的 3000 多字多属地名、人名、族名，均可知其意但不可读其音。

甲骨文在百年之后的今天已成为近代一门新兴学科——甲骨学，是历史学和古文字学的一门分支学科。它的内容包括甲骨发现历史，甲骨文流传、著录，卜辞内容研究，龟甲、兽骨种属、来源、使用部位、占卜过程、时期等。甲骨文所涉及的领域很广，包括政治、军队、战争、贡纳、农业、射猎、天文、历法、气象、建筑、疾病、祭祀等。研究甲骨文就是研究商后期的历史，用全新的手段开发、利用甲骨文资源能使初学者逐步认识甲骨文，又能使专家学者利用新的甲骨文资料获得更多得学术成果。

被誉为 20 世纪四大发现之一的甲骨文，集文献性、文物性、收藏性于一身，是考证我国商朝晚期不可多得的珍贵史料。国家图书馆珍藏甲骨 35651 片，多系名家捐赠和从市肆收购而来，约占出土总数的 1/4，是我国乃至世界上收藏甲骨最多的单位之一。

馆藏甲骨自 20 世纪 50 年代入馆珍藏以来，整理工作就没有停止过。但在甲骨自然状况最佳时期用先进的影像数据库方法整理出来，在更长久的珍藏甲骨的同时，又能让学者随时随地的通过互联网所提供的书目数据库和影像数据库，像阅读一般书籍一样对清晰的高质量彩色图像进行学术研究，应是首善之举。

一、甲骨及甲骨拓片影像数据库设计的指导思想

1. 摆脱甲骨文的流传只靠墨拓、照相、摹写这三种被沿用多年的陈旧方法。
2. 改变依据甲骨拓片进行释文、著录这一被沿用多年的研究手段。
3. 拓展从“纸上”到“纸上”考文字、谈商史的狭窄的研究空间。
4. 开辟研究甲骨的另一个半体——背面钻凿形态的新领域。
5. 更新用拓本（摹本、照片）对甲骨进行缀合的原始方法。

二、甲骨及甲骨拓片影像数据库设计的主要内容及诠释

1. 馆藏号 如：北图 19780 由于过去已被著录过的馆藏甲骨来源号均标明北图×××号，

贾双喜，国家图书馆研究馆员。

具有相对的权威性，且使用者也习惯的称为北图×××号，故馆藏号仍用北图×××号。

北图号为馆藏排架号。

2.来源号 如：罗振玉 78 善斋 25 庆云堂 24 馆藏的甲骨来源很多，其中罗伯昭（沐园）旧藏 462 片，“庆云堂”旧藏 230 片，何叙甫旧藏 71 片，孟定生旧藏 340 片，张珩旧藏 69 片，徐炳昶旧藏 24 片，曾毅公（喆厂）旧藏 181 片，郭若愚（智厂）旧藏 586 片，邵炯（茗生）旧藏 24 片，黄伯川（通古斋）旧藏 315 片，“尊古斋”旧藏 73 片，罗振玉旧藏 461 片，胡厚宣旧藏 1985 片，张仁蠡（柳风堂）旧藏 281 片，“粹雅堂”旧藏 33 片，刘体智（善斋）旧藏 28450 片，文化部转来 440 片，北图采访组购进 3 片，其他 15 片。由于过去已著录的甲骨来源号前均使用以上的人名、别号、堂号、斋号，如孟定生×××号，善斋×××号，庆云堂×××号。因此出于人们使用习惯的考虑，故设来源号。该来源号与前一项馆藏号为同一甲或同一骨。

3.贞人 贞人是甲骨卜辞分期断代的基础，据不完全统计甲骨卜辞中的贞人有一百二十多。此影像数据库依据甲骨卜辞中的贞人录入，其中包括一块甲骨多个贞人的情况也一并录入。设贞人为著录内容之一，主要的是为了弥补各家依贞人分期上的不同之说带来的不便。另外，由于个别的贞人名字库中没有，故不能按常规的方法进行检索。解决的方法，一个是进行造字，另一个是用鼠标点击影像上的贞人，以用来检索。

4.时期 殷墟甲骨文自盘庚迁殷至帝辛灭国共历 273 年，今采用董作宾先生的五个时期的分法断代，即：武丁及其以前为第一期（公元前 1300 年—公元前 1192 年，其中武丁为公元前 1250 年—公元前 1192 年），祖庚、祖甲为第二期，廪辛、康丁为第三期（因为祖庚、祖甲、廪辛、康丁共历 44 年，所以第二期、第三期的年代为公元前 1191 年—公元前 1148 年），武乙（公元前 1147 年—公元前 1113 年）、文丁（公元前 1112 年—公元前 1102 年）为第四期，帝乙（公元前 1101 年—公元前 1076 年）、帝辛（公元前 1075 年—公元前 1046 年）为第五期。

5.出土地点 坑位是甲骨卜辞断代研究标准之一。坑位一词，郭沫若先生曾说过：“坑位一项，尤非发掘者不能为”。由于馆藏甲骨的前收藏人不是甲骨的直接发掘者，故无考古学中的“层位”和“灰坑”的记载。所以在手头资料不全的情况下，只能设出土地点一项，并笼统的标明出土地点为“河南小屯”。

6.原骨属性 所指的是馆藏甲骨实物是龟甲还是兽骨。

7.原骨尺寸 原骨尺寸计量单位为厘米（cm），习惯的称为高、宽，高宽均以骨的最阔处计算。

8.卜辞内容类别 沿用《甲骨文合集》的阶级和国家、社会生产、思想文化、其它 4 大类，22 小类的分类方法。即：一. 阶级和国家。含奴隶和平民，奴隶主贵族，官吏，军队，刑罚，监狱，战争，方域，贡纳。二. 社会生产。含农业，渔猎，畜牧，手工业，商业交通。三. 思想文化。含天文历法，气象，建筑，疾病，生育，鬼神崇拜，祭祀，吉凶梦幻，卜法，文字，四、 其它。由于馆藏甲骨正处于传拓和整理工作阶段，除部分《甲骨文合集》上著录的可以参照录入外，其余的甲骨还没做此项工作。

9.拓片来源 此影像数据库的大部分拓片来源均选用本馆金石组自拓本，极少部分拓片因字迹不清晰选用以上各藏家拓本。

10.拓片尺寸 拓片尺寸计量单位为厘米（cm），习惯的称为高、宽，高宽均以拓片的最阔处计算。

11.缀合情况 主要录入根据馆藏甲骨实物进行缀合的甲骨和一些著录书籍中有关甲骨缀合的情况。

12.著录情况 指馆藏甲骨的著录情况。目前馆藏甲骨在《甲骨文合集》上著录 8000 余号，《殷契粹编》上著录 1595 号。均以《甲骨文合集》××××号，《殷契粹编》××××号的方法录入。

13.其它 指以上各项未能包括的甲骨方面特征。

三、甲骨及甲骨拓片影像数据库的实验

1. 对影像的要求

对甲骨进行高清晰度的拍摄（400dpi），使甲骨的面背文字（尤其是小字）及背面的钻凿实况展现出来。

对甲骨的特殊部位如有字的骨白，采用多角度拍摄加虚拟空间技术，要求影像旋转起来，进而看到不同的面。

对小片的甲骨只拍摄骨的正反面，附拓片参考，置于一个画面中。

2.对检索点的要求

检索点的主要途径是：馆藏号、来源号、贞人、年代、原骨属性、卜辞内容类别等。

3.对系列工具库的挂接要求

《甲骨文合集》来源表及释文部分、《甲骨文字典》、《金文字典》等。

以上甲骨及甲骨拓片影像数据库的设计和实验，是在中国数字图书馆设计和要求的各项数字化标准和原数据的整体框架中进行的。目前已做出实验性数据 50 条，并可随时查阅。