推广工程数字资源联合建设
地方文献数字化项目建设方案解读

国家图书馆　数字资源部

数字化QQ群：368907080

地方文献数字化项目工作流程

- 文献目录确认
- 数字化加工
- 数据验收

地方文献数字化项目工作流程

文献目录确认

地方文献数字化加工

- 重点建设1949年以来的地方文献，精选体现本地区地域特色，有较高文献价值和历史价值的精品馆藏资源。每个馆按照申报数量进行数字化加工。

2015重点内容

方志

- 综合志，指以某一地域范围，分类记述一定时期的政治、经济、文化的发展及其现状的志书，如《福州市志》、《大连市志》、《浙江省鄞县通志》等;
- 专志，指专门记述本地域内某一事物或某一事业的历史与现状的著述，包括专门志、专题志、部门志、行业志等独立于综合志书之外的志书，如《天津石化通志》、《浙江省青年运动志》、《越剧志》等。

地方文史资料

- 地方史料，指专门记述某一地域或社会发展过程的文献，包括地方社会史、地方经济史、地方文化史、地方革命史等诸方面,如《河北文史资料选辑》、《在井冈山的岁月》、《近代东北人民革命运动史》:旧民主主义革命时期:1840～1919》等;
- 地方人物资料，指地方知名人士的传记、评传、回忆录等，如《赖少其传》、《爱国港商潘以和》等。

地方文献数字化项目工作流程

文献目录确认

地方文献数字化加工

- 重点建设1949年以来的地方文献，精选体现本地区地域特色，有较高文献价值和历史价值的精品馆藏资源。每个馆按照申报数量进行数字化加工。

2015遴选原则

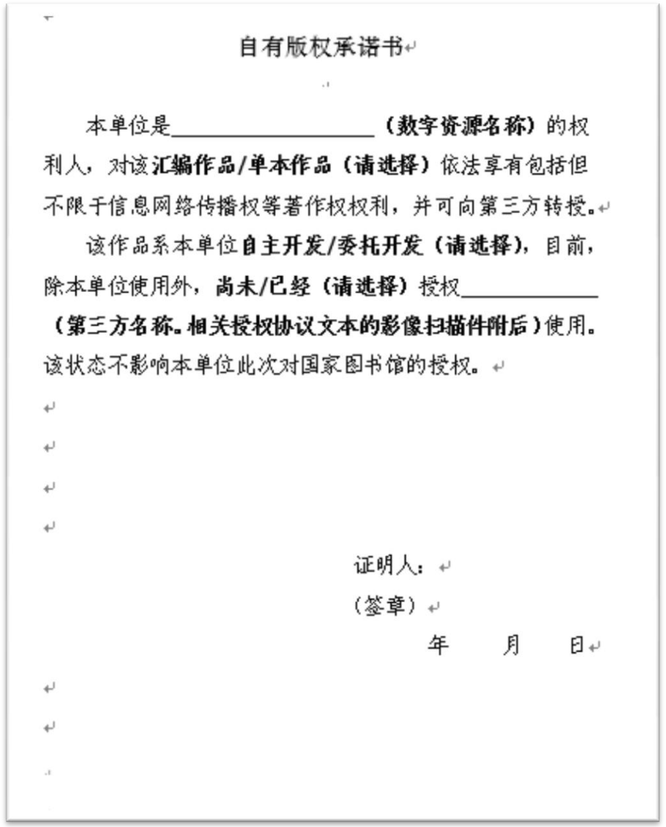
- 优先选择重点建设内容中已进入公有领域或自有版权的文献；
- 对重点建设内容按年代由远至近依次遴选；
- 本年度以汉语言文献为主，暂缓选择少数民族语言的地方文献；
- 暂缓选择地方人物的文学作品、纪念文集、研究评论文集及手稿等；
- 暂缓选择不便于OCR内容识别的地方文献，如图集、画集、歌曲集等。

地方文献数字化项目工作流程

文献目录确认

版 权 证 明

自有版权承诺书



自有版权承诺书

本单位是_____**（数字资源名称）**的权利人，对该**汇编作品/单本作品（请选择）**依法享有包括但不限于信息网络传播权等著作权权利，并可向第三方转授。

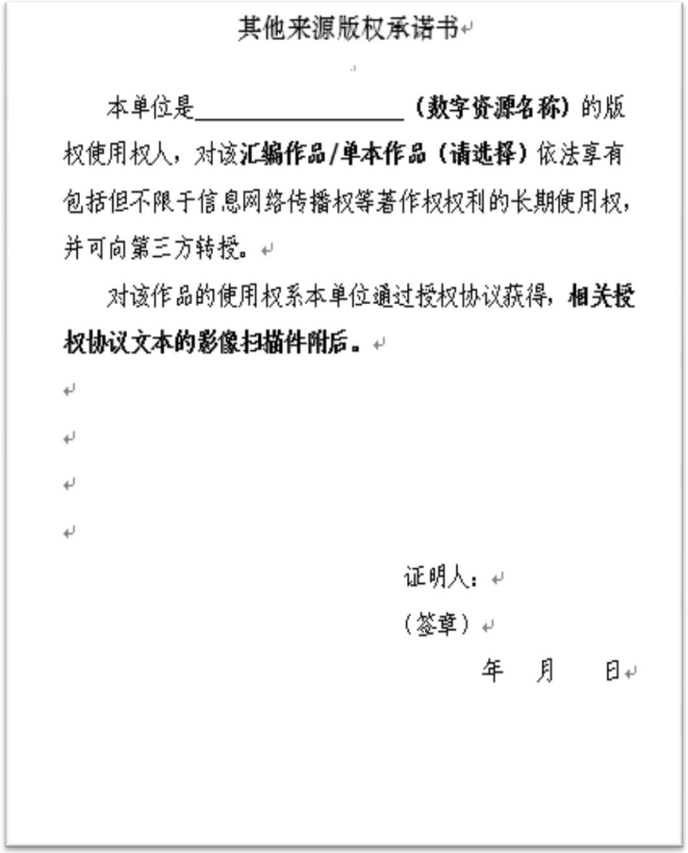
该作品系本单位**自主开发/委托开发（请选择）**，目前，除本单位使用外，**尚未/已经（请选择）**授权_____**（第三方名称。相关授权协议文本的影像扫描件附后）**使用。该状态不影响本单位此次对国家图书馆的授权。

证明人：

（签章）

年　　月　　日

其他来源版权承诺书



其他来源版权承诺书

本单位是_____**（数字资源名称）**的版权使用权人，对该**汇编作品/单本作品（请选择）**依法享有包括但不限于信息网络传播权等著作权权利的长期使用权，并可向第三方转授。

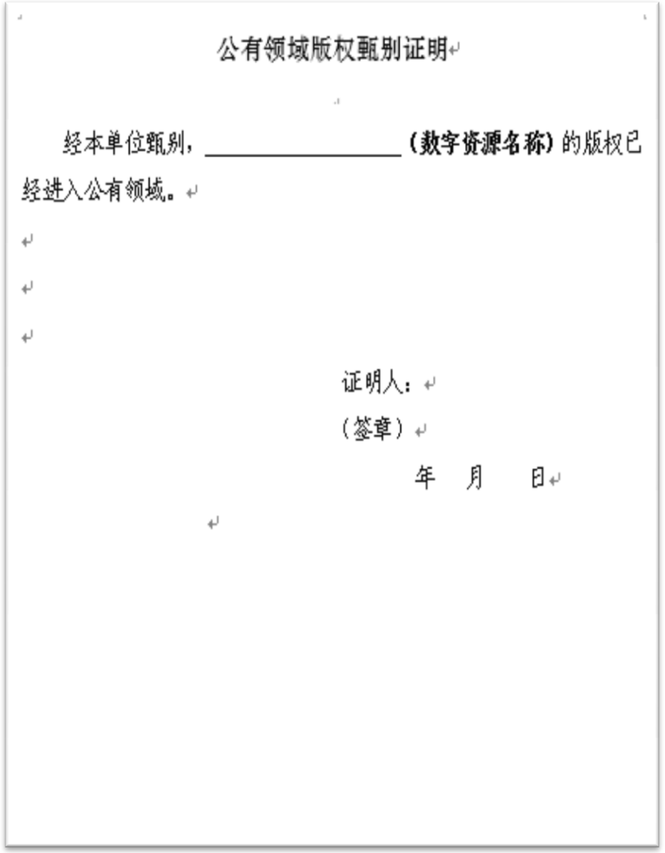
对该作品的使用权系本单位通过授权协议获得，**相关授权协议文本的影像扫描件附后。**

证明人：

（签章）

年　月　日

公有领域版权甄别证明



公有领域版权甄别证明

经本单位甄别，_____**（数字资源名称）**的版权已经进入公有领域。

证明人：

（签章）

年　月　日

地方文献数字化项目工作流程

文献目录确认

文献选取

时间：1949年之后

内容：地方特色文献包括不限于地方志，行业志。

版权：公有领域或者自有版权优先。

文献详单提交

文献详单审核

返回确认书

地方文献数字化项目工作流程

- 文献目录确认
- 数字化加工
- 数据验收

地方文献数字化项目工作流程

数字化加工

文献扫描

- 黑白页用灰度方式扫描。

- 彩色页用彩色方式扫描。

灰度方式扫描

色彩位深：8 位

分辨率：300 dpi；小于5号字体用400 dpi

档案典藏级格式：TIFF 不压缩

发布服务级格式：PDF（经过JPEG2000压缩后，再做格式转换）

彩色方式扫描

色彩位深：24 位

分辨率：300 dpi；小于5号字体用400 dpi

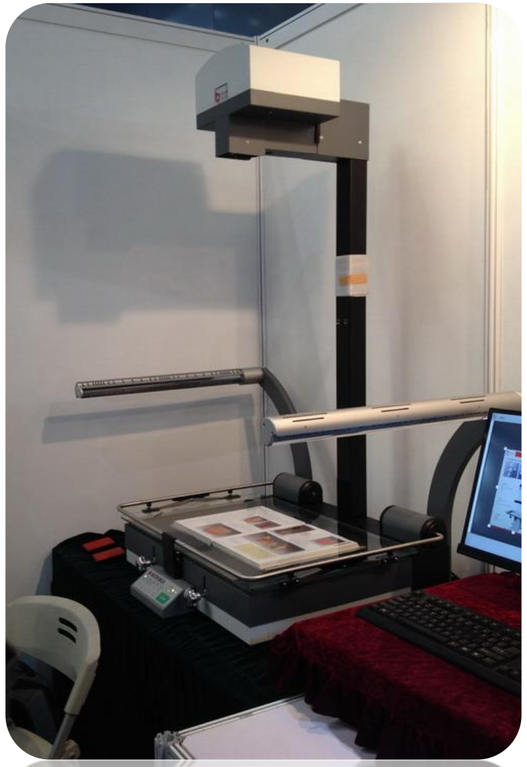
档案典藏级格式：TIFF 不压缩

发布服务级格式：PDF（经过JPEG2000压缩后，再做格式转换）

地方文献数字化项目工作流程

数字化加工

高清扫描仪



高速扫描仪

零边距扫描仪

地方文献数字化项目工作流程

数字化加工

图像质量要求

数字化环境注意防护光源，避免透光或反射光的影响。数字化后的图像清晰，文件页码连续，没有重页、缺页，错页等情况（原书缺页、错页除外）。补扫的图像要与同册图像文件的大小一致，颜色接近。

1.以原文献的上边沿为基准，以中缝为中心线，保持原文献的天头、地脚的尺寸不变，左右两边的尺寸基本不变。
2.数字图像放大至实际尺寸100%，图像不失真。
3.数字图像文件与文献原件颜色不一致，须先进行设备色彩校正，再重新进行扫描或拍照工作。

地方文献数字化项目工作流程

数字化加工

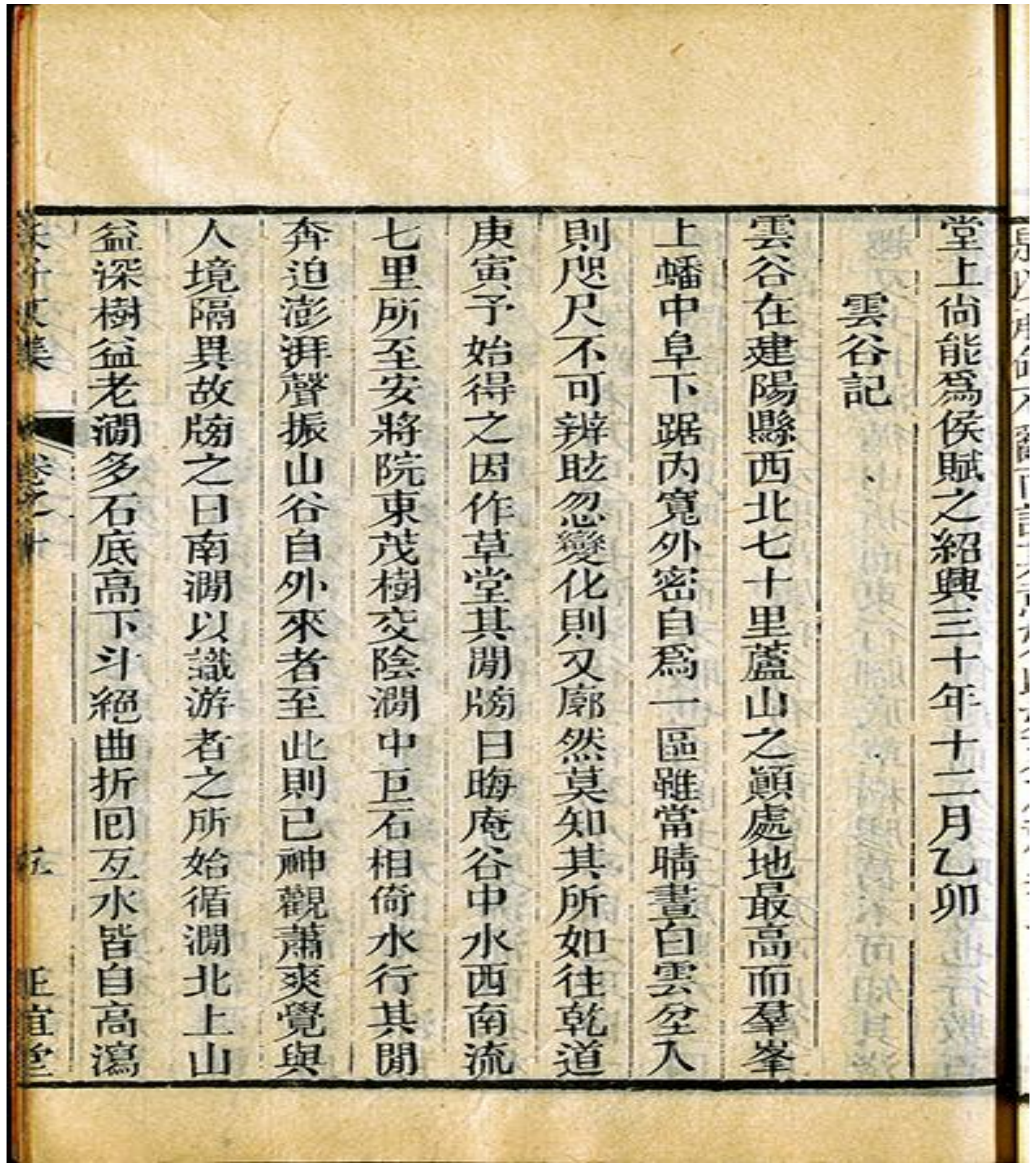
图片反射光源

无反射情况

地方文献数字化项目工作流程

数字化加工

透字问题

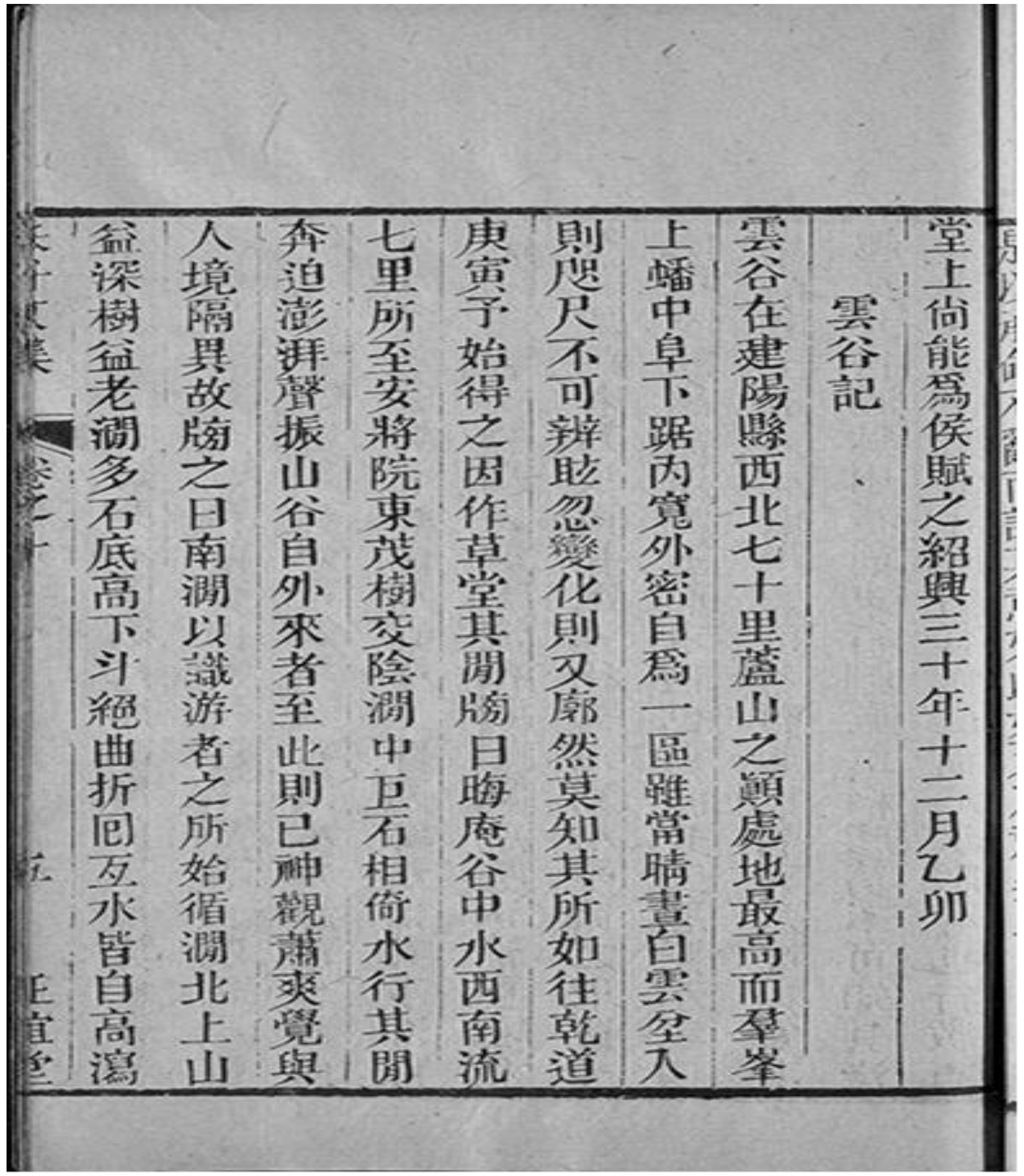


堂上倘能爲侯賦之紹興三十年十二月乙卯

雲谷記

雲谷在建陽縣西北七十里蘆山之巔處地最高而羣峯上蟠中阜下踞內寬外密自爲一區雖當晴晝白雲坌入則咫尺不可辨眩忽變化則又廓然莫知其所如往乾道庚寅予始得之因作草堂其閒牓曰晦庵谷中水西南流七里所至安將院東茂樹交陰澗中巨石相倚水行其閒奔迫澎湃聲振山谷自外來者至此則已神觀蕭爽覺與人境隔異故牓之曰南澗以識游者之所始循澗北上山益深樹益老澗多石底高下斗絕曲折回互水皆自高瀉

无透字状态



堂上倘能爲侯賦之紹興三十年十二月乙卯

雲谷記

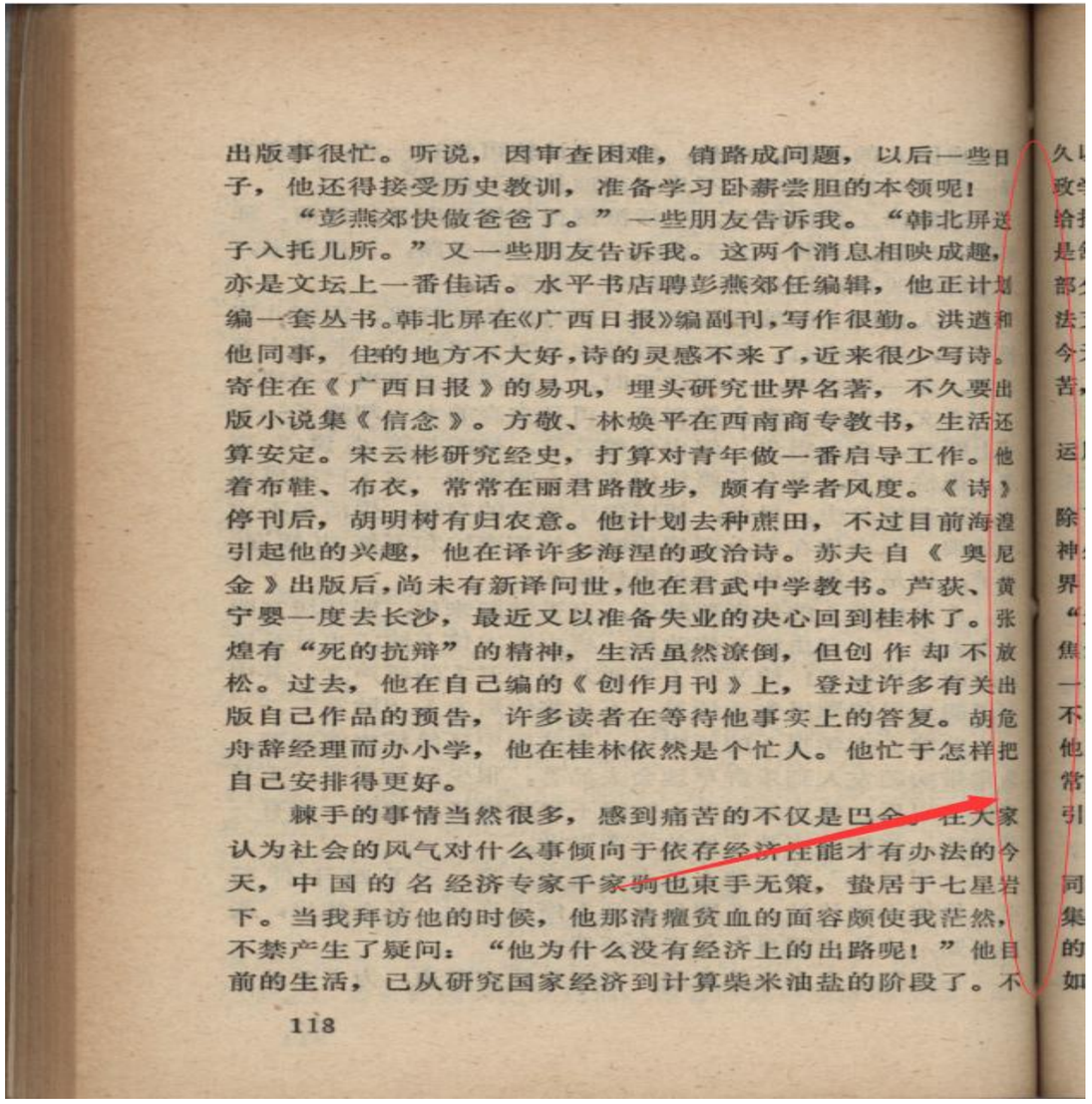
雲谷在建陽縣西北七十里蘆山之巔處地最高而羣峯上蟠中阜下踞內寬外密自爲一區雖當晴晝白雲坌入則咫尺不可辨眩忽變化則又廓然莫知其所如往乾道庚寅予始得之因作草堂其閒牓曰晦庵谷中水西南流七里所至安將院東茂樹交陰澗中巨石相倚水行其閒奔迫澎湃聲振山谷自外來者至此則已神觀蕭爽覺與人境隔異故牓之曰南澗以識游者之所始循澗北上山益深樹益老澗多石底高下斗絕曲折回互水皆自高瀉

地方文献数字化项目工作流程

数字化加工

文字偏斜，变形

边距不当



出版事很忙。听说，因审查困难，销路成问题，以后一些日子，他还得接受历史教训，准备学习卧薪尝胆的本领呢！

“彭燕郊快做爸爸了。”一些朋友告诉我。“韩北屏送子入托儿所。”又一些朋友告诉我。这两个消息相映成趣，亦是文坛上一番佳话。水平书店聘彭燕郊任编辑，他正计划编一套丛书。韩北屏在《广西日报》编副刊，写作很勤。洪遒和他同事，住的地方不大好，诗的灵感不来了，近来很少写诗。寄住在《广西日报》的易巩，埋头研究世界名著，不久要出版小说集《信念》。方敬、林焕平在西南商专教书，生活还算安定。宋云彬研究经史，打算对青年做一番启导工作。他着布鞋、布衣，常常在丽君路散步，颇有学者风度。《诗》停刊后，胡明树有归农意。他计划去种蔗田，不过目前海涅引起他的兴趣，他在译许多海涅的政治诗。苏夫自《奥尼金》出版后，尚未有新译问世，他在君武中学教书。芦荻、黄宁婴一度去长沙，最近又以准备失业的决心回到桂林了。张煌有“死的抗辩”的精神，生活虽然潦倒，但创作却不放松。过去，他在自己编的《创作月刊》上，登过许多有关出版自己作品的预告，许多读者在等待他事实上的答复。胡危舟辞经理而办小学，他在桂林依然是个忙人。他忙于怎样把自己安排得更好。

棘手的事情当然很多，感到痛苦的不仅是巴金。在大家认为社会的风气对什么事倾向于依存经济性能才有办法的今天，中国的名经济专家千家驹也束手无策，蛰居于七星岩下。当我拜访他的时候，他那清癯贫血的面容颇使我茫然，不禁产生了疑问：“他为什么没有经济上的出路呢！”他目前的生活，已从研究国家经济到计算柴米油盐的阶段了。不

118

地方文献数字化项目工作流程

数字化加工

TIFF图像

佛教传入瑞安始于何时，现已很难考定。因为佛教最初传入只有少数人信仰，未必为上层官府和史官所注意。据佛教经籍记载：汉明帝时，两湖、江、浙一带佛教比中原更加兴盛，汉明帝的弟弟刘英信奉佛教可以证明这一点。至于佛教在汉明帝前曾经已有印度高僧来瑞安传教，民间有少数人信奉的说法，还没有确实的证据。汉灵帝建宁年中（170）佛经汉译创始人——安息国太子安息高，自称宿世尚欠有命债在会稽未还，特地从广州东游到会稽寻找宿世冤家债主，当时瑞安属会稽郡地，安息高在中国时达三十年之久，他在会稽期间到瑞安传播佛法，这是非常有可能的。

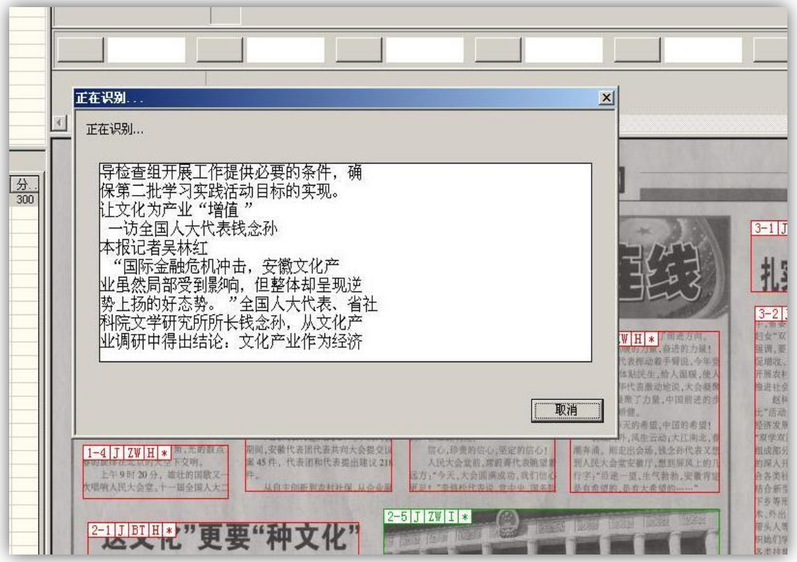
佛教在瑞安得到政府承认崇信，并初步建立了它的基础和规模是始于南朝永初年间（420），当时永嘉太守谢灵运笃信佛法，并皈依慧远法师，对庐山东林寺净土初祖慧远及刘遗民等十八高贤十分敬慕。慧远在东林寺创办净土道场，谢灵运在东林寺东西开凿二池，种上白莲，表示对佛教虔诚信仰。并作有《石室饭僧》和《净土诗》，赞扬三宝与净土法门。义熙十二年，谢灵运为慧远撰文立碑。

太守信佛，佛教肯定已在民间流传，梁天监二年（503），瑞安东镇街兴建东安寺，陶山泮岱谢岙龙山山麓建聚福寺，继而西岘山麓建栖霞寺（后改名悟真寺今改建人民剧院），场桥山麓建龙翔寺。自此，瑞安佛教逐渐得到传播。

1

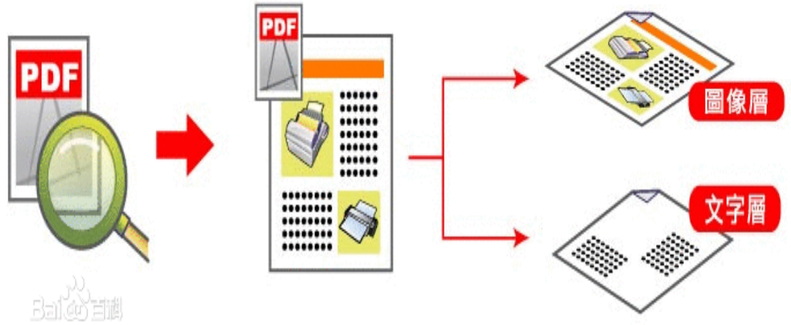
TIFF图像均为单页

TXT文件



TXT文件有单版和合并版

双层PDF

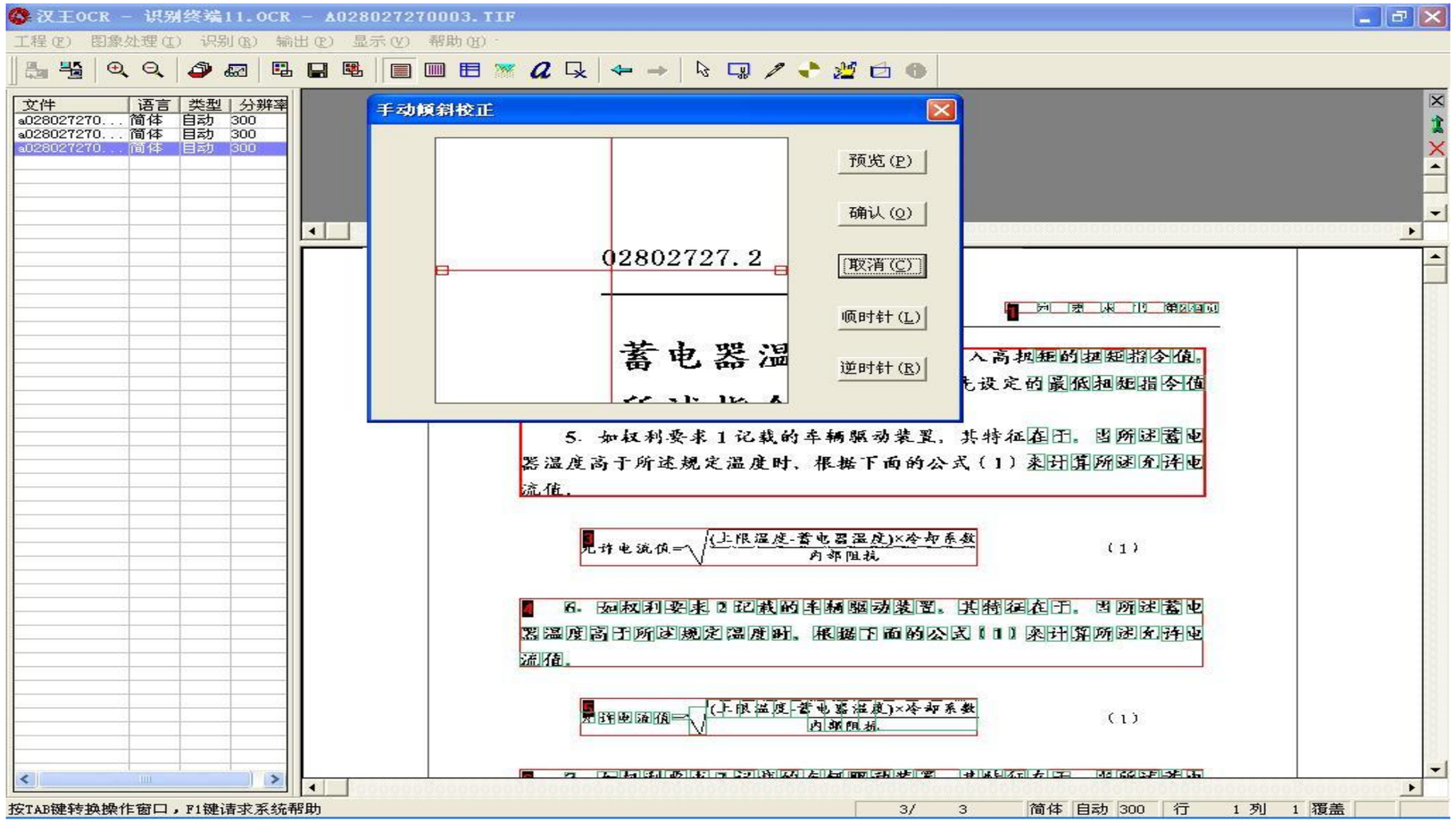


双层PDF有单版和合并版

地方文献数字化项目工作流程

数字化加工

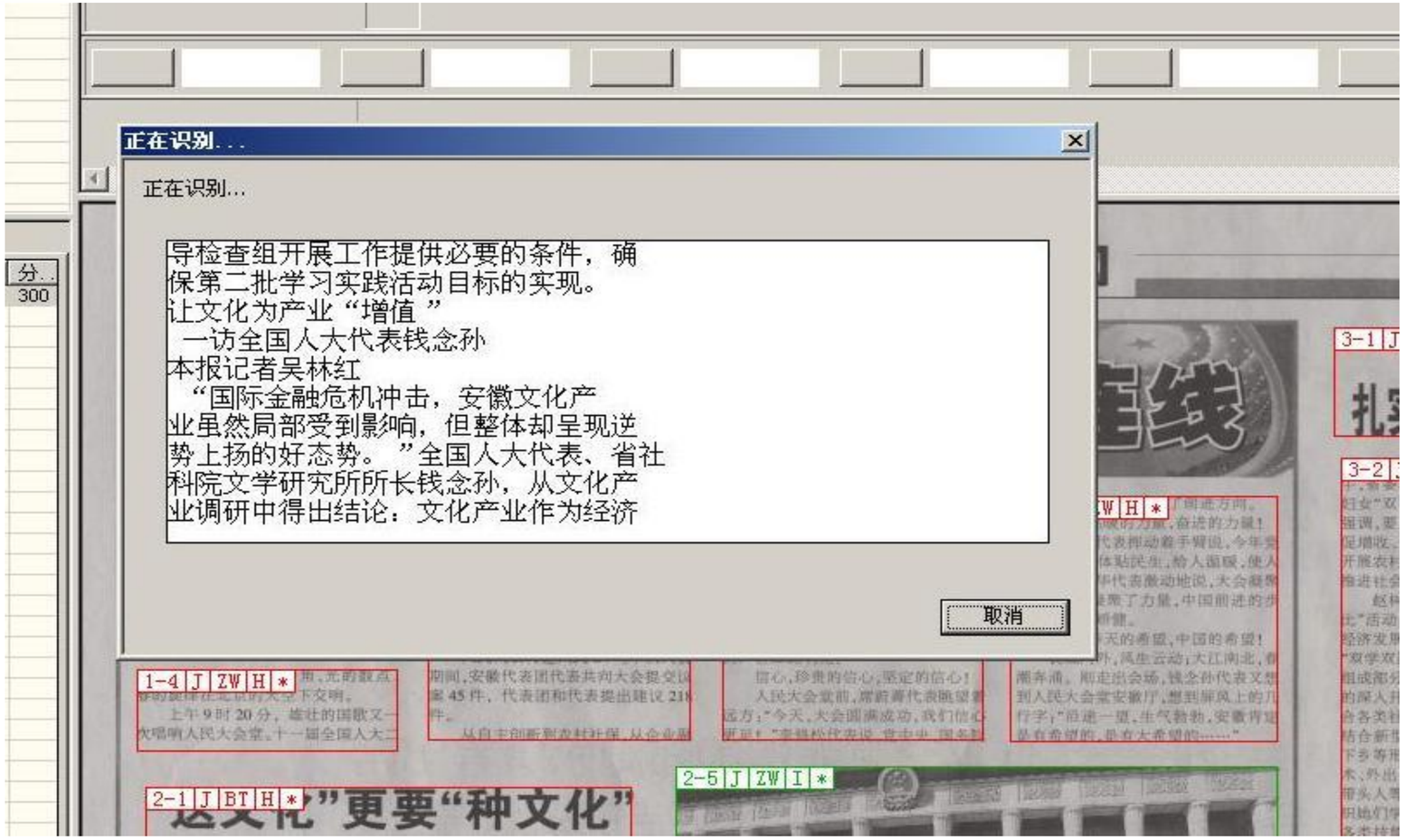
版面分析



地方文献数字化项目工作流程

数字化加工

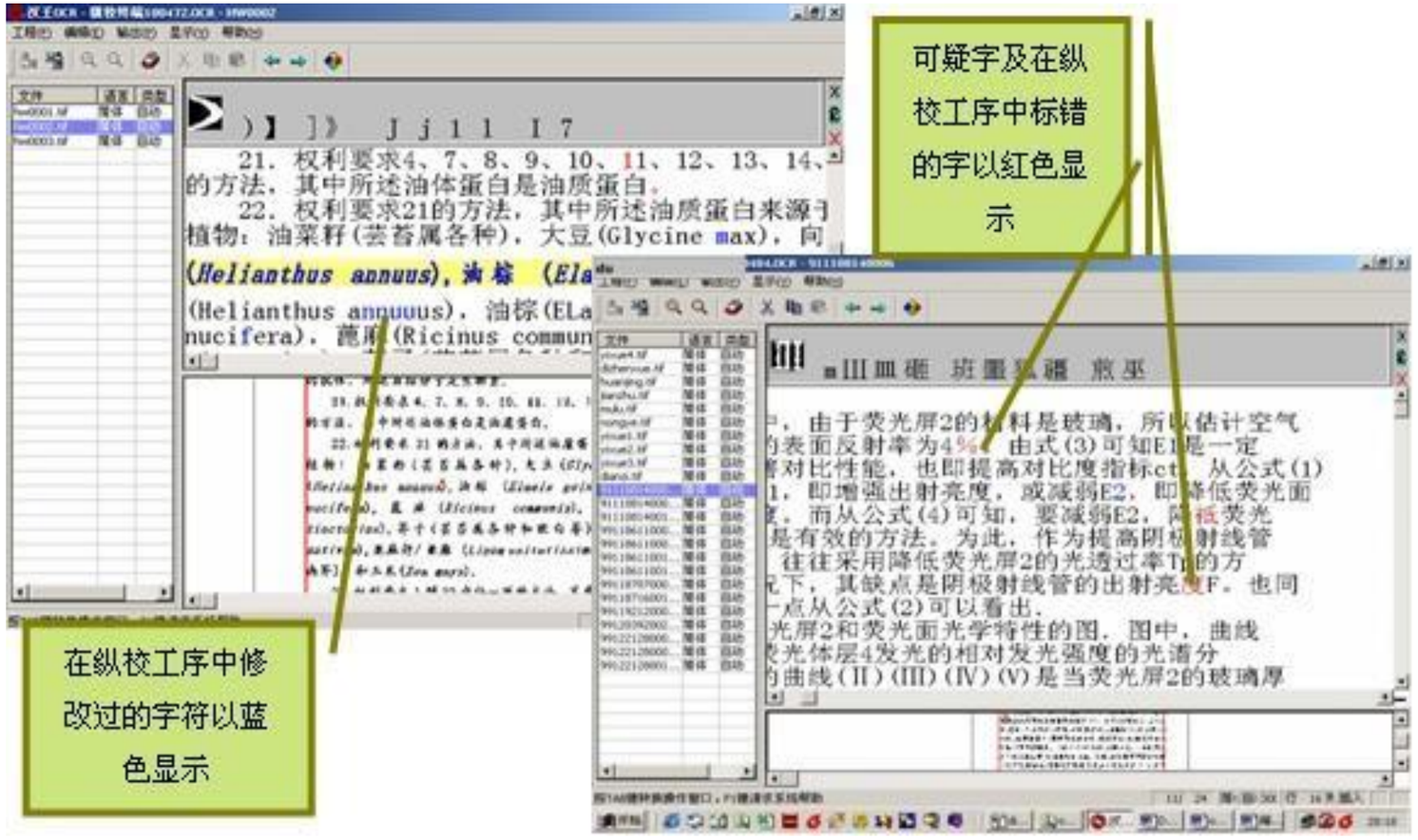
自动识别



地方文献数字化项目工作流程

数字化加工

字符校对



地方文献数字化项目工作流程

数字化加工

文本质量要求

1.文本数据应如实反映原书内容、版面等所有原书相关信息。
2.文本数据保存格式为TXT纯文本格式。
3.文件命名无误，且在数量上与TIFF图像一致。
4.文本数据内容与TIFF图像内容吻合，不存在乱码、转换错误等问题。
5.文本数据应如实反映原文的章节、段落，不应出现与文章段落不符的字符、段落、硬回车、空格等。
6.单版TXT文件与合并版TXT文件内容完全一致。

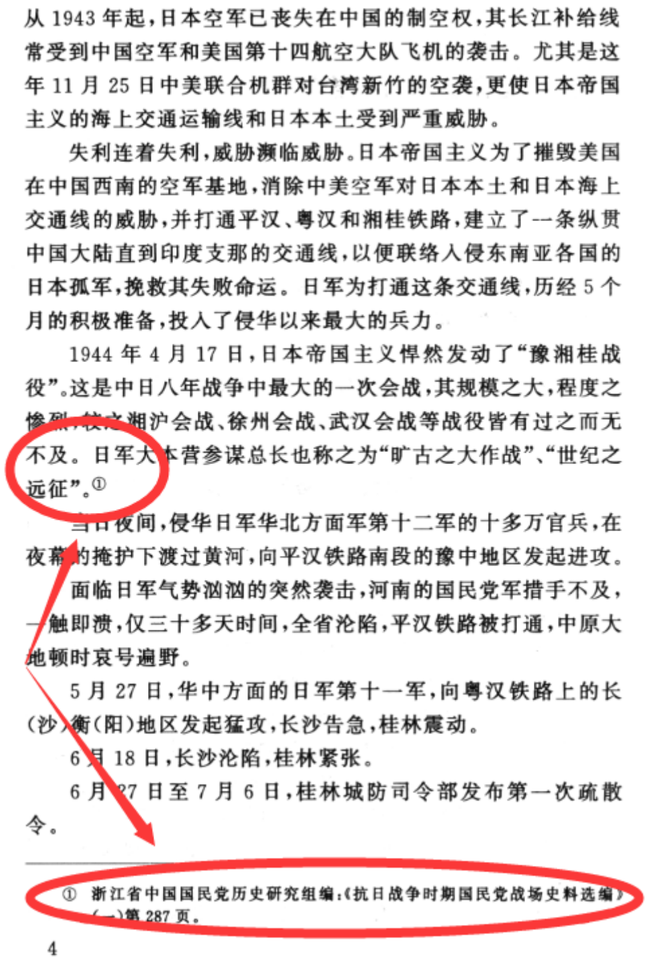
特殊情况的处理

注释

注释分注释类和解说类，仅对有意义注释进行转换，无意义注释不做转换。

- **注释出现在当页，且明确标出与正文对应位置的：**将注释内容填入括号“（）”内，插回到原文注释所对应的位置。
- **注释在一篇文章或章节结尾，以参考或引用等专项标题单独列出的：**按原文版式转换，不插回原文。
- **解说类注释出现在正文文字段落中间或左右两侧，在正文中无对应位置的：**将注释内容放在其出现的那段文字后，另起一段，段首标注“注释：”。
- **解说类注释出现在黑框或深色底框内，且在正文文字段落中间的将注释内容放在【】内的：**原位置转换。

地方文献数字化工作流程



从 1943 年起，日本空军已丧失在中国的制空权，其长江补给线常受到中国空军和美国第十四航空大队飞机的袭击。尤其是这年 11 月 25 日中美联合机群对台湾新竹的空袭，更使日本帝国主义的海上交通运输线和日本本土受到严重威胁。

失利连着失利，威胁濒临威胁。日本帝国主义为了摧毁美国在中国西南的空军基地，消除中美空军对日本本土和日本海上交通线的威胁，并打通平汉、粤汉和湘桂铁路，建立了一条纵贯中国大陆直到印度支那的交通线，以便联络入侵东南亚各国的日本孤军，挽救其失败命运。日军为打通这条交通线，历经 5 个月的积极准备，投入了侵华以来最大的兵力。

1944 年 4 月 17 日，日本帝国主义悍然发动了“豫湘桂战役”。这是中日八年战争中最大的一次会战，其规模之大，程度之惨烈，较之湘沪会战、徐州会战、武汉会战等战役皆有过之而无不及。日军大本营参谋总长也称之为“旷古之大作战”、“世纪之远征”。①

当日夜间，侵华日军华北方面军第十二军的十多万官兵，在夜幕的掩护下渡过黄河，向平汉铁路南段的豫中地区发起进攻。

面临日军气势汹汹的突然袭击，河南的国民党军措手不及，一触即溃，仅三十多天时间，全省沦陷，平汉铁路被打通，中原大地顿时哀号遍野。

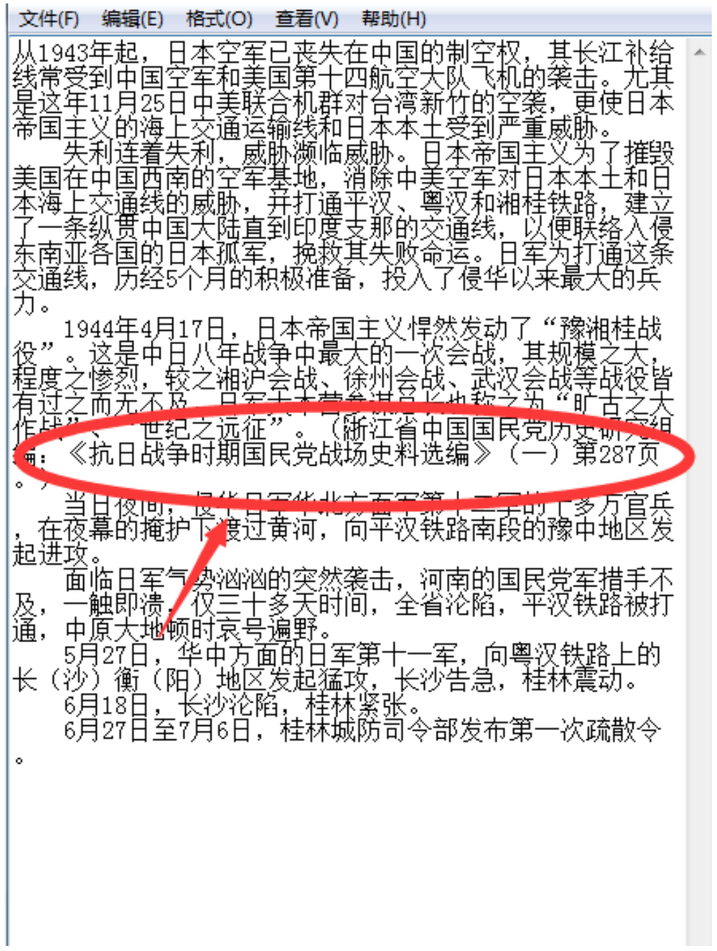
5 月 27 日，华中方面的日军第十一军，向粤汉铁路上的长（沙）衡（阳）地区发起猛攻，长沙告急，桂林震动。

6 月 18 日，长沙沦陷，桂林紧张。

6 月 27 日至 7 月 6 日，桂林城防司令部发布第一次疏散令。

① 浙江省中国国民党历史研究组编：《抗日战争时期国民党战场史料选编》（一）第 287 页。

4



文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

从1943年起，日本空军已丧失在中国的制空权，其长江补给线常受到中国空军和美国第十四航空大队飞机的袭击。尤其是这年11月25日中美联合机群对台湾新竹的空袭，更使日本帝国主义的海上交通运输线和日本本土受到严重威胁。

失利连着失利，威胁濒临威胁。日本帝国主义为了摧毁美国在中国西南的空军基地，消除中美空军对日本本土和日本海上交通线的威胁，并打通平汉、粤汉和湘桂铁路，建立了一条纵贯中国大陆直到印度支那的交通线，以便联络入侵东南亚各国的日本孤军，挽救其失败命运。日军为打通这条交通线，历经5个月的积极准备，投入了侵华以来最大的兵力。

1944年4月17日，日本帝国主义悍然发动了“豫湘桂战役”。这是中日八年战争中最大的一次会战，其规模之大，程度之惨烈，较之湘沪会战、徐州会战、武汉会战等战役皆有过之而无不及。日军大本营参谋总长也称之为“旷古之大作战”、“世纪之远征”。（浙江省中国国民党历史研究组编：《抗日战争时期国民党战场史料选编》（一）第287页。）

当日夜间，侵华日军华北方面军第十二军的十多万官兵，在夜幕的掩护下渡过黄河，向平汉铁路南段的豫中地区发起进攻。

面临日军气势汹汹的突然袭击，河南的国民党军措手不及，一触即溃，仅三十多天时间，全省沦陷，平汉铁路被打通，中原大地顿时哀号遍野。

5月27日，华中方面的日军第十一军，向粤汉铁路上的长（沙）衡（阳）地区发起猛攻，长沙告急，桂林震动。

6月18日，长沙沦陷，桂林紧张。

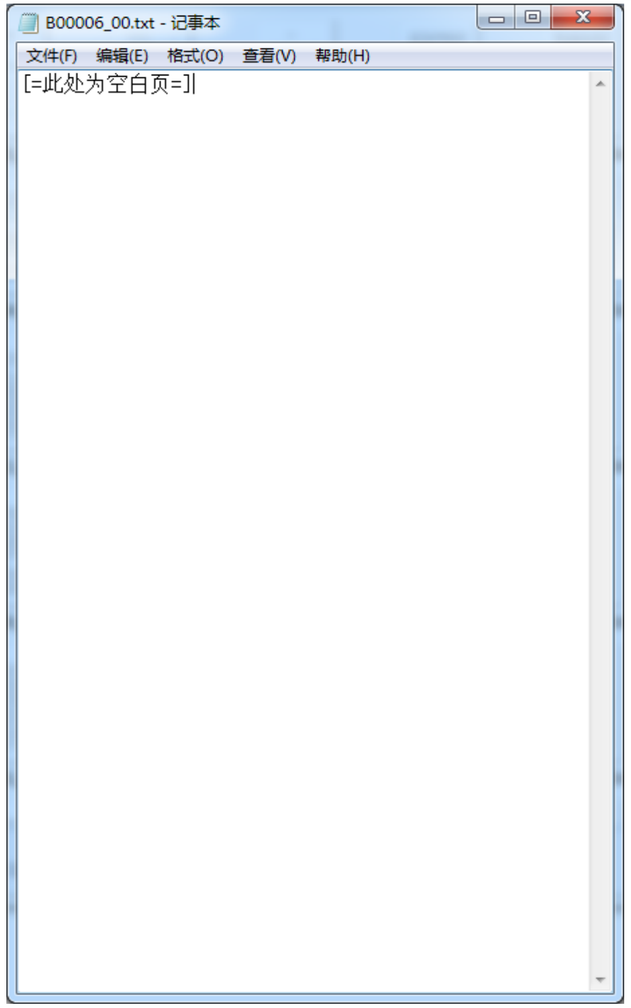
6月27日至7月6日，桂林城防司令部发布第一次疏散令。

注释出现在当页，且明确标出与正文对应位置的：将注释内容填入括号“（）”内，插回到原文注释所对应的位置。

特殊情况的处理

空白页

为保持内容完整性和页面连贯性，正文中空白页需保留，并按照命名规则正确命名，内容标注为“[=此处为空白页=]”。



特殊情况的处理

插图与插图页

- 插图

仅对有意义插图进行转换和标注。在插图出现的段落后另起一段，标注为“[=此处为插图（图注）=]”。

- 插图页

均需保留，并按照命名规则正确命名，内容标注为：“[=此处为插图页（图注）=]”；对于包含多个插图的插图页，内容标注为：“[=此处为插图页：图一（图注），图二（图注）...=]”。

地方文献数字化项目工作流程

但從父輩們對他為人的稱讚中，他那可敬佩的形象，早已在我們的心目中形成，並常以他那種百折不撓、自強不息的奮鬥精神，來鼓勵自己奮發向前。我們都是六十年代初畢業的大學生，本應該在自己創造性的領域裏發揮才華，為社會作出貢獻。可是不久一場「史無前例的文化大革命」來臨，到了「五七幹校」，只是勞動鍛煉。那時日，我們都沒有忘記在勞動之餘，對自己的專業做點研究，珍惜難得的每一寸光陰，才有現在的一點小小成就！

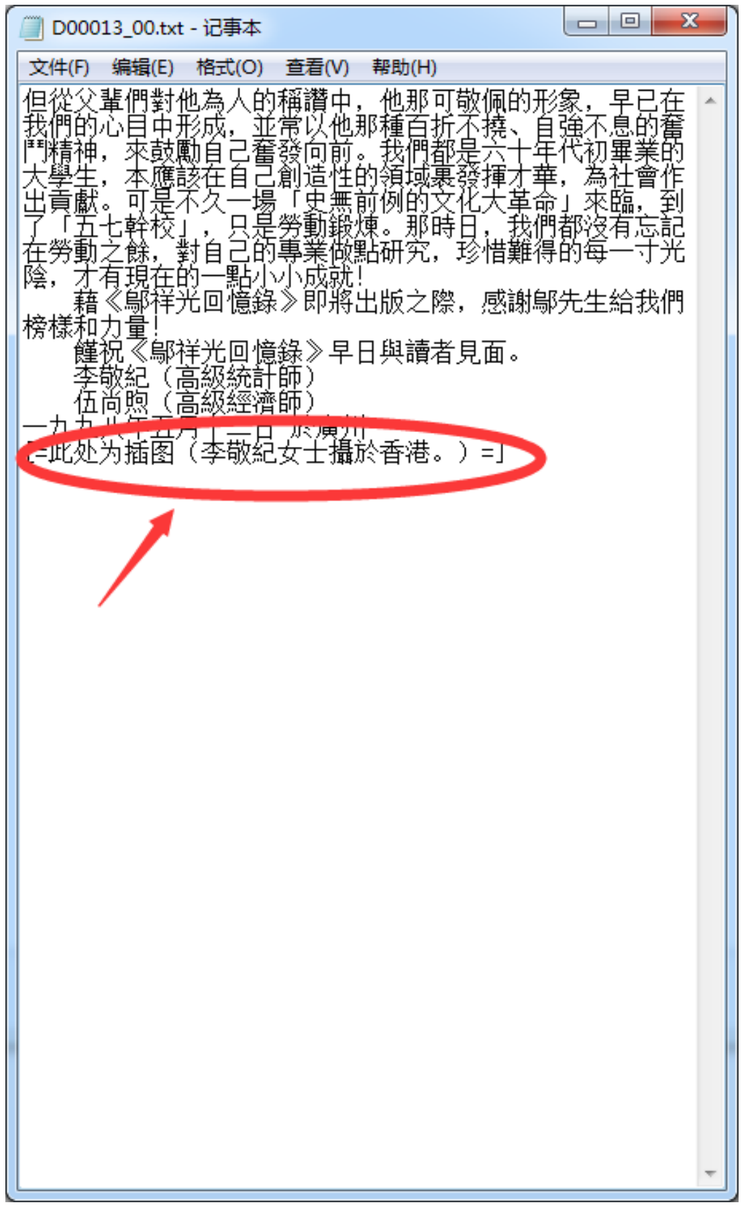
藉《鄔祥光回憶錄》即將出版之際，感謝鄔先生給我們榜樣和力量！

謹祝《鄔祥光回憶錄》早日與讀者見面。

李敬紀（高級統計師）
伍尚煦（高級經濟師）

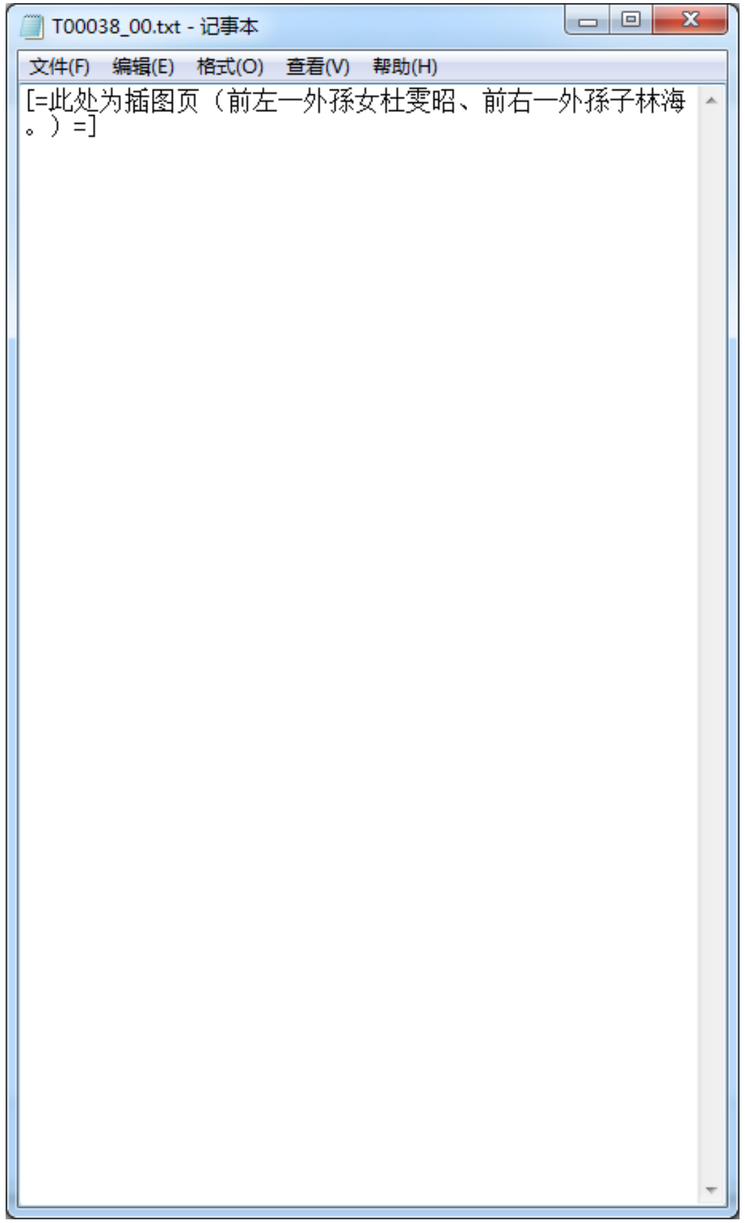
一九九八年五月十二日　於廣州

李敬紀女士攝於香港。



地方文献数字化项目工作流程

前左一外孙女杜雯昭、前右一外孙子林海。



特殊情况的处理

表格与表格页

表格只转换表格内文字及表注，任何形式表格边框不予转换。

- 表格

简单的单列或横向排版表格应予转换；多列复杂表格，内容为分类列举的文字按列予以转换；其他表格不予转换，在表格出现的位置，标注为“[=此处为表格（表格说明）=]”。

- 表格页

需保留，并按照命名规则正确命名，内容标注为“[=此处为表格页（表格说明）=]”。

对未进行转换的图像和表格，需在其所属单板TXT文件目录内建立“未转换文件对应图像”子目录，将此插图/表格所在的图像文件保存在该子目录内；多个插图/表格对应同一个图像时，仅保存一个图像文件。同时对未进行转换的插图/表格在《文献全文转换未转换文件记录表》中进行记录。

地方文献数字化项目工作流程

数字化加工

简单表格

一、国内林氏组织

名　称	名　称
中国卫辉比干庙文化研究会	广东南雄市西河堂林氏宗亲会
河南省林氏总会	江西省资溪县林氏研究会
河南省比干历史文献研究会	海南比干学述研究会
河南新乡市姓氏文化研究会	山东济南比干学述研究会
河南省鹤壁市朝歌林氏亲宗会	山东陶县比干学述研究会
福建省姓氏源流研究会	浙江省杭州市林氏研究会
福建省林氏源流研究委员会	浙江省泰顺县林氏研究会
福建省晋江市比干学述研究会	浙江省温州市林氏联合会
福建省晋江市金井学述研究会	浙江省苍南市林氏联合会
福建省莆田林氏始祖文物管理董事会	浙江省平阳林氏宗亲会
福建省莆田湄州岛妈祖研究会	浙江省永嘉林氏宗亲会
福建省莆田九牧林氏祖祠董事会	浙江省瑞安林氏宗亲会
闽林始祖文物保护研究会	浙江省温岭林氏宗亲会
闽湄州妈祖庙董事会	浙江省乐清林氏宗亲会
闽南安殷武柴比干学述研究会	浙江省临海林氏研究会
闽泉州武荣殷比干文化研究会	浙江省青田林氏研究会
福建省福鼎市林氏研究会	浙江省巨州市林氏研究会
石狮市比干文化研究会	浙江省永康市林氏研究会
中国黔滇林氏研究会	台湾林氏宗亲会
广西宾阳比干学述研究会	台湾台中林氏宗庙
广西桂林比干学术研究会	香港中华林西河堂联合总会
广西桂花比干学术研究会	香港新令石咀同乡会
广西合浦林氏家庙	香港连滩林氏宗亲会
广东陆丰林氏研究会	香港陶江林氏族亲会
广东潮州林氏研究会	

复杂表格

2006年末乡镇(街道)行政村居民区分布

乡镇(街道)名称	行政村数	居民社区数	乡镇(街道)名称	行政村数	居民社区数
瞻岐镇	17	0	钟公庙街道	25	17
咸祥镇	17	1	高桥镇	20	3
塘溪镇	17	0	横街镇	28	2
东钱湖镇	41	5	集士港镇	19	1
东吴镇	12	1	古林镇	24	3
五乡镇	19	2	石碶街道	16	3
邱隘镇	17	3	洞桥镇	20	1
下应街道	25	5	鄞江镇	12	1
云龙镇	17	1	龙观乡	10	0
横溪镇	15	1	章水镇	20	1
姜山镇	55	4	梅墟街道	12	2
全区合计	458	57			

特殊情况的处理

拼音文字混编

拼音文字混编分为对全文均做拼音标注和仅对个别文字做拼音标注两种情况。

- 全文均做拼音标注的（拼音一般标注在文字上方）

只转换文字，不转换拼音

- 个别文字做拼音标注的（拼音一般标注在文字后面）

需按照原文版式，同时转换文字和拼音。

特殊情况的处理

下划线

- 下划线标记的为单个字符或字母的

  将标注内容填入【】，放在有下划线的字符或字母后。

- 下划线标注的为一句话或一段字符或字母的

  将有下划线的内容放在（）内，标注内容填入【】，放在有下划线的内容后。

There is a tree　转换后　There is【A】 a tree↵

A↵

一句话或一段字符或字母：↵

There is a tree　转换后　（There is a tree）【B】↵

B↵

特殊情况的处理

其他

无法录入的生僻字、公式、符号等内容用“〓”表示。同时将“〓”所对应图像文件保存在单版TXT文件目录内建立“未转换文件对应图像”文件夹内。

文件夹建立方法：

- 子目录内应包含所有用“〓”表示的图像文件并以jpg格式保存。
- 图像文件删除其他信息，仅保留“无法录入的内容”信息，图像命名不变。
- 多个用“〓”表示的内容对应同一个图像时，仅保存一个图像文件。

地方文献数字化项目工作流程

“音乐是比一切智慧与哲学更高的启示”。在写作这件作品时，他又说：“从我的心里流出来，流到大众的心里。”

全曲依照弥撒祭曲礼的程序，（按弥撒祭歌唱的词句，皆有经文——拉丁文的——规定，任何人不能更易一字，各段文字大同小异，而节目繁多，谱为音乐时部门尤为庞杂。凡不解经典及不知典礼的人较难领会。）分成五大颂曲：（一）吾主怜我 *Kyrie*；（二）荣耀归主 *Gloria*；（三）我信我主 *Credo*；（四）圣哉圣哉 *Sanctus*；（五）神之羔羊 *Agnus Dei*（全曲以四部独唱与管弦乐队及大风琴演出。乐队的构成如下：2 flûtes；2 hautbois；2 clarinettes；2 bassons；1 contrebasse；4 cors（horns）；2 trompettes；2 trombones；timbale 外加弦乐五重奏，人数之少非今人想象所及。）——第一部以热诚的祈祷开始，继以 Andante 奏出“怜我怜我”的悲叹之声，对基督的呼吁，在各部合唱上轮流唱出（五大部每部皆如奏鸣曲式分成数章，兹不详解）。——第二部表示人类俯伏卑恭，颂赞上帝，歌颂主荣，感谢恩赐。——第三部，贝多芬流露出独有的口吻了。开始时的庄严巨大的主题，表现他坚决的信心。结实的节奏，特殊的色彩，trompette 的运用，作者把全部乐器的机能用来证实他的意念。他的神是胜利的英雄，是半世纪后尼采所宣扬的“力”的神。贝多芬在耶稣的苦难上发现了他自身的苦难。在受难、下葬等壮烈悲哀的曲调以后，接着是复活的呼声，英雄的神明胜利了！——第四部，贝多芬参见了神明，从天国回到人间，散布一片温柔的情绪。然后如《第九交响曲》一般，是欢乐与轻快的爆发。紧接着祈祷，苍茫的，神秘的。虔诚的信徒匍匐着，已经蒙到

“音乐是比一切智慧与哲学更高的启示”。在写作这件作品时，他又说：“从我的心里流出来，流到大众的心里。”

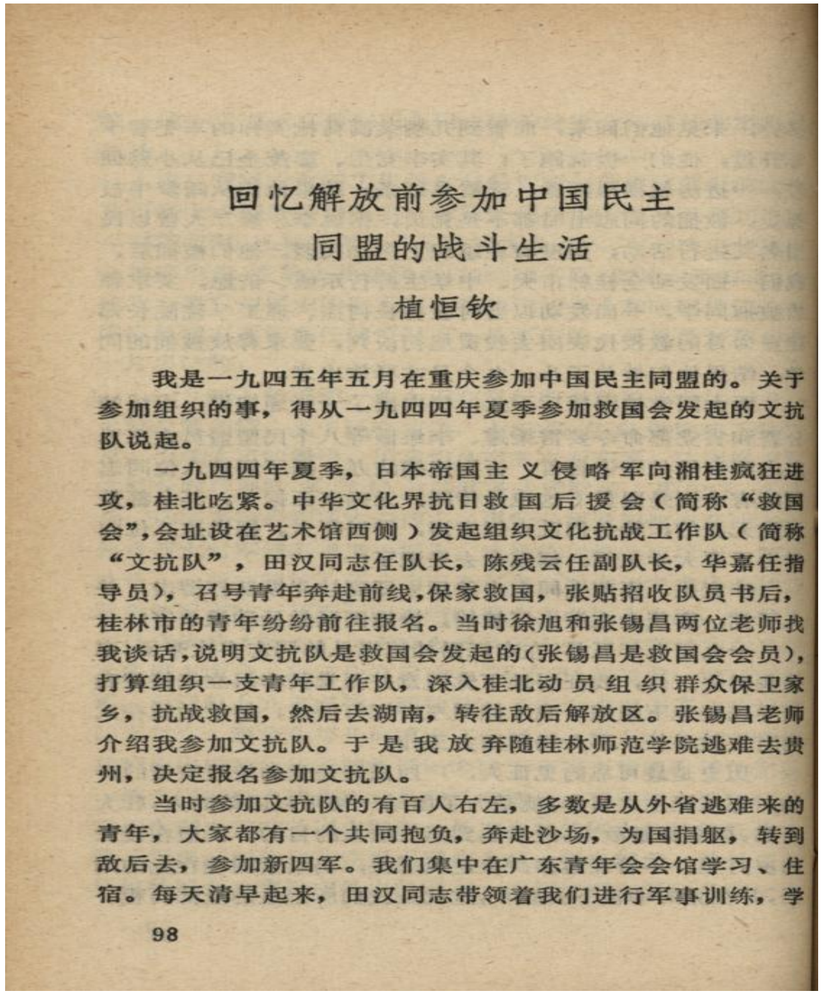
全曲依照弥撒祭曲礼的程序，（按弥撒祭歌唱的词句，皆有经文——拉丁文的——规定，任何人不能更易一字，各段文字大同小异，而节目繁多，谱为音乐时部门尤为庞杂。凡不解经典及不知典礼的人较难领会。）分成五大颂曲：（一）吾主怜我Kyrie；（二）荣耀归主Gloria；（三）我信我主Credo；（四）圣哉圣哉Sanctus；（五）神之羔羊Agnus Dei（全曲以四部独唱与管弦乐队及大风琴演出。乐队的构成如下：2 flûtes；2 hautbois；2 clarinettes；2 bassons；1 contrebasse；4 cors（horns）；2 trompettes；2 trombones；timbale外加弦乐五重奏，人数之少非今人想象所及。）——第一部以热诚的祈祷开始，继以Andante奏出“怜我怜我”的悲叹之声，对基督的呼吁，在各部合唱上轮流唱出（五大部每部皆如奏鸣曲式分成数章，兹不详解）。——第二部表示人类俯伏卑恭，颂赞上帝，歌颂主荣，感谢恩赐。——第三部，贝多芬流露出独有的口吻了。开始时的庄严巨大的主题，表现他坚决的信心。结实的节奏，特殊的色彩，trompette的运用，作者把全部乐器的机能用来证实他的意念。他的神是胜利的英雄，是半世纪后尼采所宣扬的“力”的神。贝多芬在耶稣的苦难上发现了他自身的苦难。在受难、下葬等壮烈悲哀的曲调以后，接着是复活的呼声，英雄的神明胜利了！——第四部，贝多芬参见了神明，从天国回到人间，散布一片温柔的情绪。然后如《第九交响曲》一般，是欢乐与轻快的爆发。紧接着祈祷，苍茫的，神秘的。虔诚的信徒匍匐着，已经蒙到

地方文献数字化项目工作流程

数字化加工

文本常见问题

原文



回忆解放前参加中国民主
同盟的战斗生活

植恒钦

我是一九四五年五月在重庆参加中国民主同盟的。关于参加组织的事，得从一九四四年夏季参加救国会发起的文抗队说起。

一九四四年夏季，日本帝国主义侵略军向湘桂疯狂进攻，桂北吃紧。中华文化界抗日救国后援会（简称“救国会”，会址设在艺术馆西侧）发起组织文化抗战工作队（简称“文抗队”，田汉同志任队长，陈残云任副队长，华嘉任指导员），召号青年奔赴前线，保家救国，张贴招收队员书后，桂林市的青年纷纷前往报名。当时徐旭和张锡昌两位老师找我谈话，说明文抗队是救国会发起的（张锡昌是救国会会员），打算组织一支青年工作队，深入桂北动员组织群众保卫家乡，抗战救国，然后去湖南，转往敌后解放区。张锡昌老师介绍我参加文抗队。于是我放弃随桂林师范学院逃难去贵州，决定报名参加文抗队。

当时参加文抗队的有百人右左，多数是从外省逃难来的青年，大家都有一个共同抱负，奔赴沙场，为国捐躯，转到敌后去，参加新四军。我们集中在广东青年会会馆学习、住宿。每天清早起来，田汉同志带领着我们进行军事训练，学

98

转换后

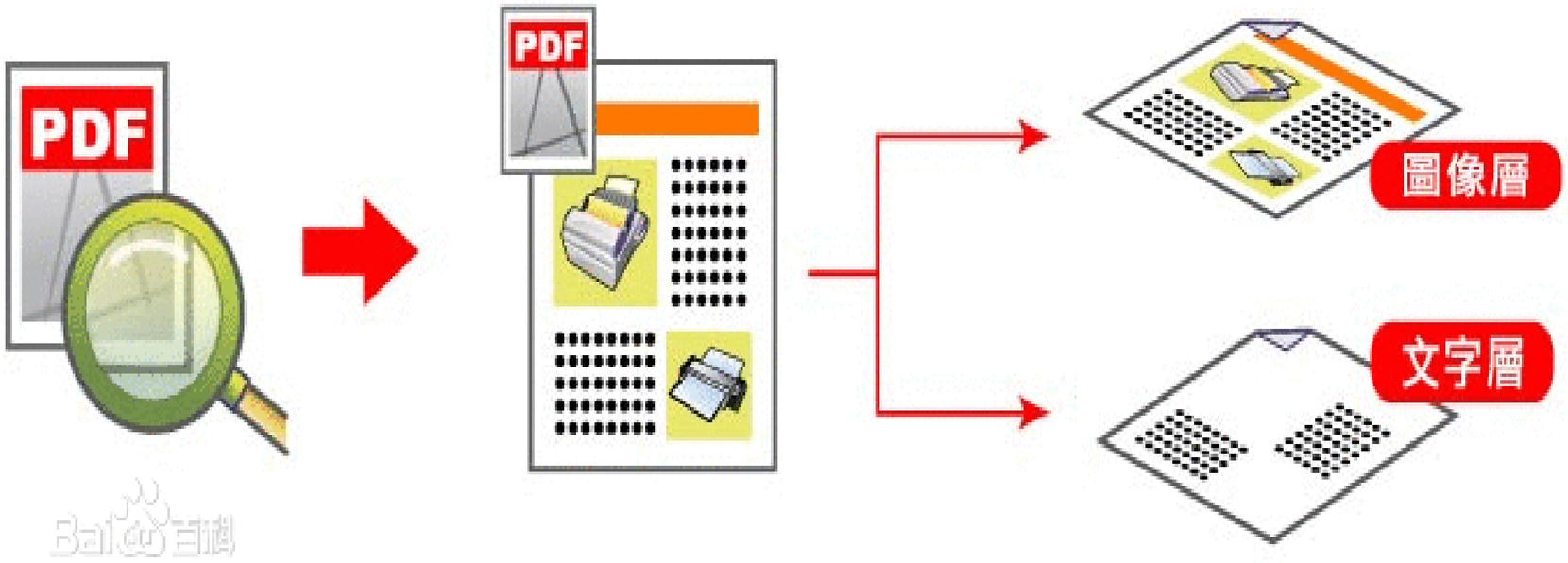
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

回忆解放前参加中国民主 同盟的战斗生活
植恒钦
我是一九四五年五月在重庆参加中国民主同盟的。关于 参加组织的事，得从一九四四年夏季参加救国会发起的文抗 队说起。
一九四四年夏季，日本帝国王义侵略军向湘桂疯狂进 攻，桂北吃紧。中华化界抗日救国后援会（简称“救国 会”，会址设在艺术馆西侧）发起组织文化抗战工作队（简称 “文抗队”，田汉同志任队长，陈残云任副队长，华嘉任指 导员），召号青年奔赴前线，保家救国，张贴招收队员书后， 桂林市的青年纷纷前往报名。当时徐旭和张锡昌两位老师找 我谈话，说明文抗队是救国会发起的(张锡昌是救国会会员)， 打算组织一支青年工作队，深入桂北动员组织群众保卫家 乡，抗战救国，然后去湖南，转往敌后解放区。张锡昌老师 介绍我参加文抗队。于是我放弃随桂林师范学院逃难去贵 州，决定报名参加文抗队。
当时参加文抗队的有百人右左，多数是从外省逃难来的 青年，大家都有一个共同抱负，奔赴沙场，为国捐躯，转到 敌后去，参加新四军。我们集中在广东青年会会馆学习、住 宿。每天清早起来，田汉同志带领着我们进行军事训练，学
98

地方文献数字化项目工作流程

数字化加工

双层PDF



地方文献数字化项目工作流程

数字化加工

PDF质量要求

- 双层PDF数据需完整，避免缺页、重页、页码顺序颠倒等问题。
- 双层PDF文件的图像层和文字层的文字对位准确，反显区域与文字区域相差1毫米以内。
- 打开一本电子文献阅览并对文字放大时，保证在放大到百分之二百的时候，字迹清晰，笔画连续，无断裂、缺块的现象。
- 整本PDF必须制作书签。书签是电子书的目录，内容和纸质书的目录一致。书签的功能是用户点击书签的某一章节，电子书会自动显示为相应页。打开PDF文件时，自动显示书签，书签只展开到第一级目录。

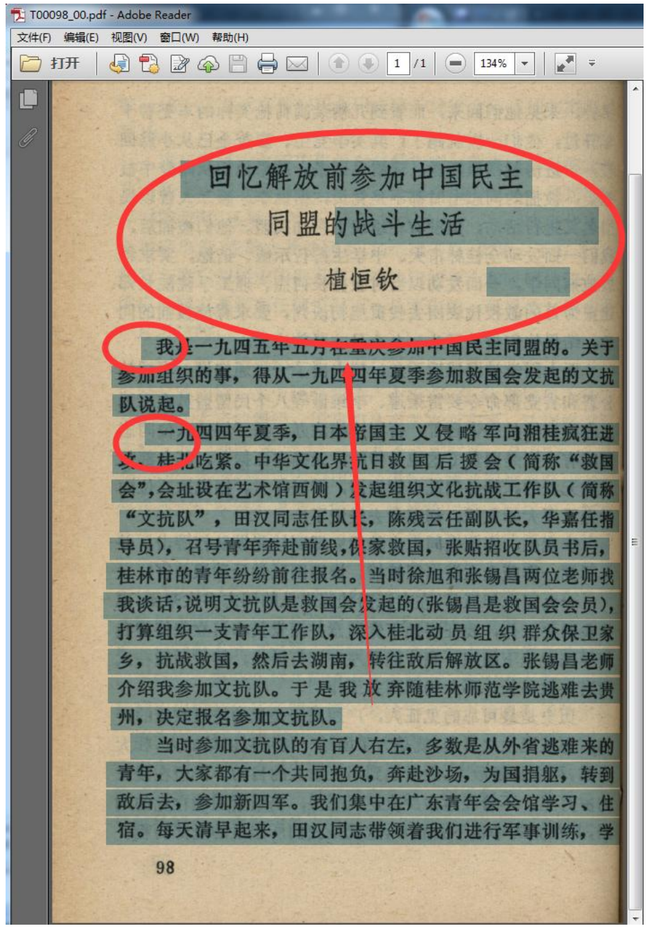
地方文献数字化项目工作流程

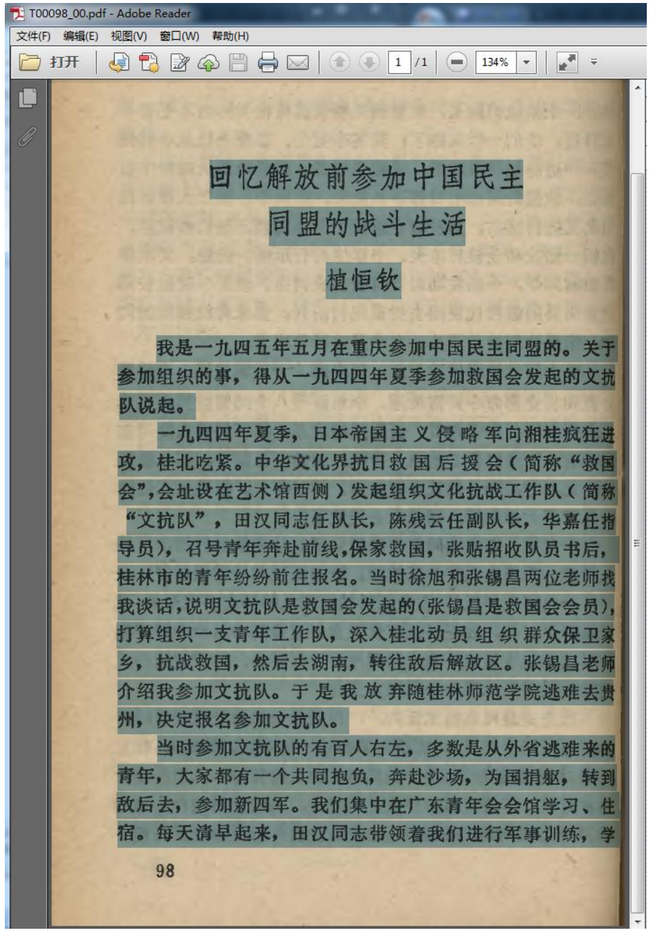
数字化加工

PDF常见问题

图像层

文字层





地方文献数字化项目工作流程

数字化加工

对象命名

加工编号：文献数字化加工过程中一册文献的唯一标识，11位数字和1位下划线

第 1 位：文献基本资料类型，图书为0

第 2 位：文献语种，中文为1

第 3-4 位：加工年，公元年后两位数字

第 5-8 位：机构代码，前2位表示省，后两位表示市，黑龙江0800

第 9-11位：单位内部流水号，自行分配，从1开始，不足3位以0补齐。

1位下划线在机构代码之后。

01150800_001	黑龙江省志·总述
01150800_002	黑龙江省志·大事记
01150800_003	黑龙江省志·地理志
01150800_004	黑龙江省志·地质矿产志
01150800_005	黑龙江省志·气象志地震志
01150800_006	黑龙江省志·经济综志
01150800_007	黑龙江省志·农业志
01150800_008	黑龙江省志·土地志
01150800_009	黑龙江省志·畜牧志
01150800_010	黑龙江省志·水产志
01150800_011	黑龙江省志·林业志
01150800_012	黑龙江省志·农机志
01150800_013	黑龙江省志·国营农场志

地方文献数字化项目工作流程

数字化加工

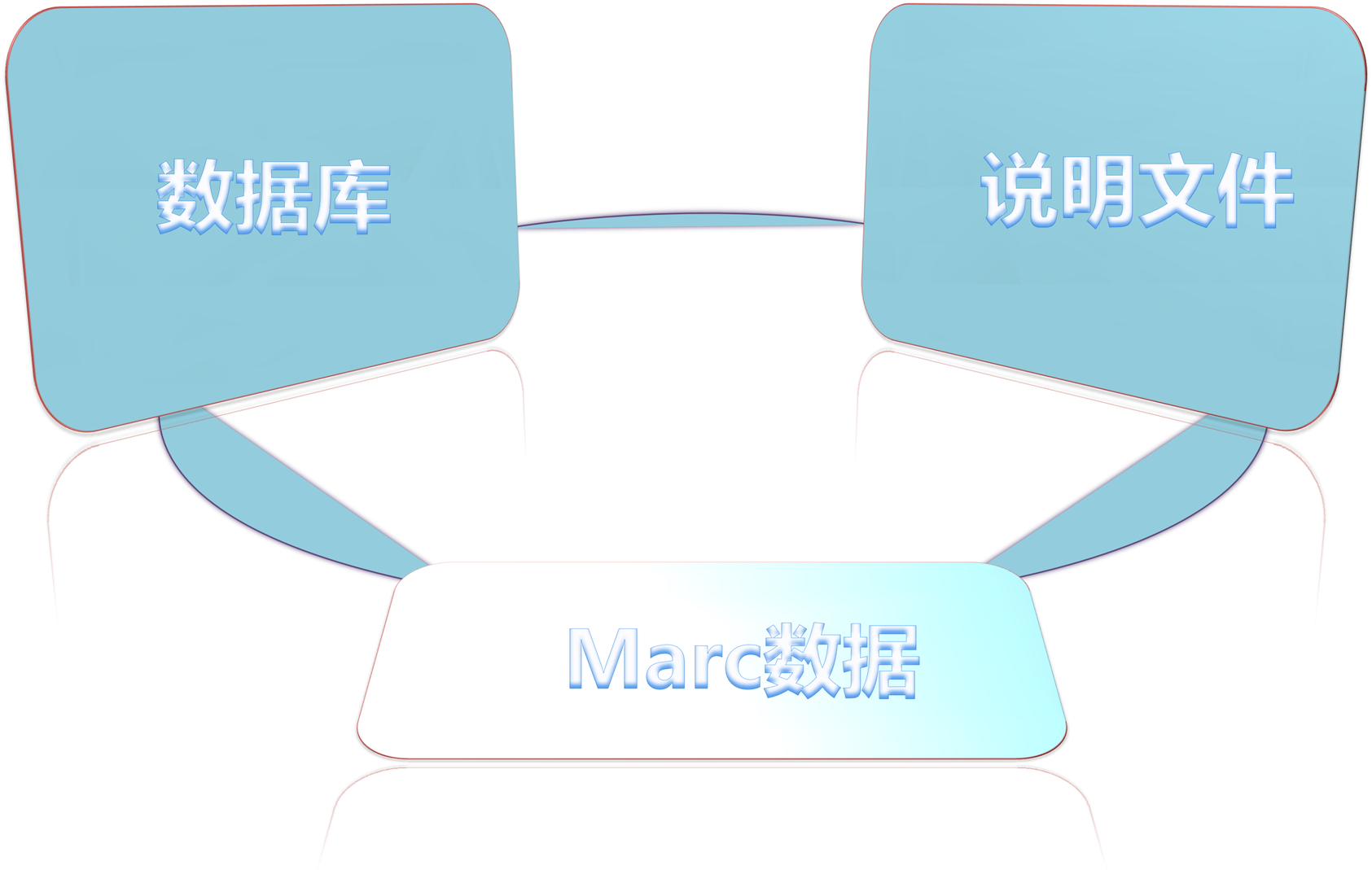
对象命名

- **前封（含封一、封二）**
  - 扫描文件名为Axxxxx_00，其中xxxxx为5位数字，按原书顺序依次排序
- **前附页**
  - 目录页之前的前附页扫描文件名为Bxxxxx_00，其中xxxxx为5位数字，按原书顺序依次排序。
  - 目录页之后的前附页扫描文件名为Dxxxxx_00，其中xxxxx为5位数字，按原书顺序依次排序
- **目录页**
  - 扫描文件名为Cxxxxx_00，其中xxxxx为5位数字，按原书顺序依次排序。
- **正文**
  - 有页码的正文扫描文件名为Txxxxx_00,其中xxxxx为5位数字,与原书页号一致，按原书顺序依次排序。
  - 正文中插页扫描文件名为Txxxxx_vv，其中xxxxx为5位数字,表示插页的前一页顺序号,vv为数字，表示插页，并按原书顺序依次排序。
- **后附页**
  - 扫描文件名为Yxxxxx_00，其中 xxxxx 为5位数字，按原书顺序依次排序。
- **后封（含封三、封四）**
  - 扫描文件名为Zxxxxx_00，其中xxxxx为5位数字,按原书顺序依次排序。

地方文献数字化项目工作流程

数字化加工

相关文档



地方文献数字化项目工作流程

数字化加工

数据库

数据库命名

8位：

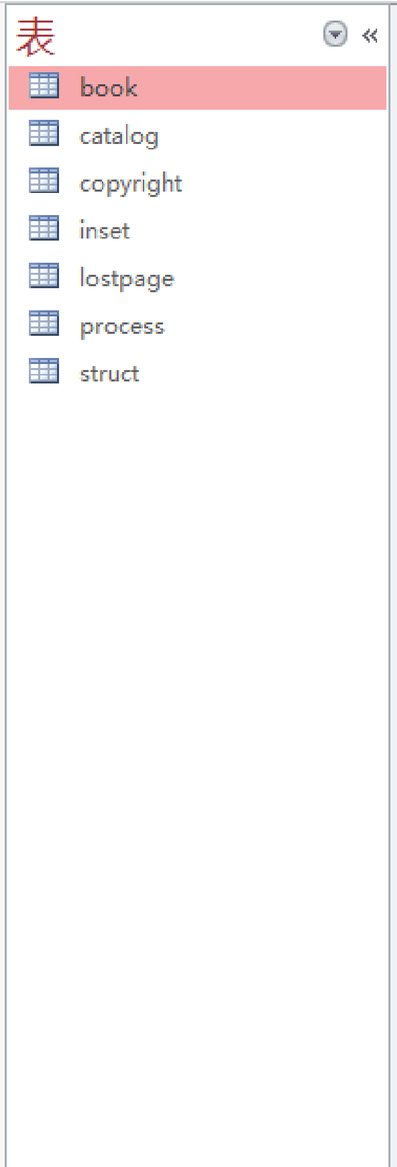
文献基本资料类型（1位）

文献语种（1位）

加工年（2位）

机构代码（4位）。

如：01140100.mdb



book表

book									
book_id	cat_id	book_name	author	pub_house	pub_date	isbn	record_id	barcode	cdoi
01140100_001	K835.165.76=41	贝多芬传	(法)罗曼·罗兰著	华文出版社	2013	978-7-5075-3899-1	006517927	3293936997	

序号	中文名称	字段名称	对应书目数据（MARC）
1	加工编号	book_id	
2	分类	cat_id	第一个690字段$a
3	书名	book_name	200字段$a.$h,$i,$e
4	作者	author	200字段$f
5	出版社	pub_house	210字段$c
6	出版时间	pub_date	210字段$d
7	ISBN号	isbn	010字段$a
8	001	record_id	001字段
9	条码号	barcode	
10	唯一标识符	cdoi	

样例

book - Access

表格工具

文件 开始 创建 外部数据 数据库工具 字段 表

视图 粘贴 剪切 复制 格式刷 筛选器 升序 降序 取消排序 选择 高级 切换筛选 全部刷新 新建 保存 删除 合计 拼写检查 其他 查找 替换 转至 选择 调整至窗体大小 切换窗口 宋体 11 B I U

视图 剪贴板 排序和筛选 记录 查找 窗口 文本格式

表

- book
- catalog
- copyright
- inset
- lostpage
- process
- struct

book_id	cat_id	book_nam	author	pub_house	pub_date	isbn	record_id	barcode	cdoi
01140100 001	K835.165.76=41	贝多芬传	(法)罗曼·罗兰著	华文出版社	2013	978-7-5075-3899-1	006517927	3293936997	
*									

表

- book
- catalog
- copyright
- inset
- lostpage
- process
- struct

catalog表

book_id	serial_nu	chapter_num	chapter_name	author	page_num	ppage_num	page_plac	page_prop
01140100_001	1	目录						1
01140100_001	2		译者序		1	1	D	0
01140100_001	3		原序		1	3	D	0
01140100_001	4		初版序		1	6	D	0
01140100_001	5		贝多芬传		5	3	T	0
01140100_001	6		贝多芬遗嘱			51	T	0
01140100_001	7		贝多芬遗嘱		53	53	T	0
01140100_001	8		海林根施塔特遗嘱		54	54	T	0
01140100_001	9		书信集			59	T	0
01140100_001	10		贝多芬致阿门达牧师书		61	61	T	0
01140100_001	11		贝多芬致弗兰茨·格哈得·韦格勒书		64	64	T	0
01140100_001	12		致韦格勒书		68	68	T	0
01140100_001	13		贝多芬致韦格勒书		75	75	T	0
01140100_001	14		致韦格勒书		77	77	T	0
01140100_001	15		贝多芬致莫舍勒斯书		78	78	T	0
01140100_001	16		思想录			79	T	0
01140100_001	17		关于音乐		81	81	T	0
01140100_001	18		关于批评		85	85	T	0
01140100_001	19		参考书目		89	89	T	0
01140100_001	20		附录			91	T	0
01140100_001	21		贝多芬评传		93	93	T	0
01140100_001	22		贝多芬的作品及其精神		110	110	T	0
01140100_001	23	一、	贝多芬与力		110	110	T	0
01140100_001	24	二、	贝多芬的音乐建树		117	117	T	0
01140100_001	25	三、	重要作品浅释		126	126	T	0
01140100_001	26		论莫扎特		150	150	T	0
01140100_001	27		编者后记		173	173	T	0

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	标引序号	serial_num	数值型
3	章节号	chapter_num	
4	章节名	chapter_name	
5	作者	author	
6	页码	page_num	客观著录，如实反映目录页原貌（可为空）
7	绝对页码	ppage_num	数值型，文件名数字部分
8	页位置	page_place	文件名字母部分
9	属性	page_prop	1）“目录”属性为“1”；2）“无目录”属性为“2”； 3）每册书除第一条目录外，其余记录的属性默认为“0”数值型

样例

Access

表格工具

文件 开始 创建 外部数据 数据库工具 字段 表

视图 剪切 复制 格式刷 粘贴 筛选器 升序 降序 取消排序 选择 高级 切换筛选 全部刷新 新建 保存 删除 合计 拼写检查 其他 查找 替换 转至 选择 调整至窗体大小 切换窗口 宋体 11 B I U

视图 剪贴板 排序和筛选 记录 查找 窗口 文本格式

表

- book
- catalog
- copyright
- inset
- lostpage
- process
- struct

catalog

book_id	serial_nu	chapter_num	chapter_name	author	page_num	ppage_num	page_plac	page_prop
01140100_001	1	目录						1
01140100_001	2		译者序		1	1	D	0
01140100_001	3		原序		1	3	D	0
01140100_001	4		初版序		1	6	D	0
01140100_001	5		贝多芬传		5	3	T	0
01140100_001	6		贝多芬遗嘱			51	T	0
01140100_001	7		贝多芬遗嘱		53	53	T	0
01140100_001	8		海林根施塔特遗嘱		54	54	T	0
01140100_001	9		书信集			59	T	0
01140100_001	10		贝多芬致阿门达牧师书		61	61	T	0
01140100_001	11		贝多芬致弗兰茨·格哈得·韦格勒书		64	64	T	0
01140100_001	12		致韦格勒书		68	68	T	0
01140100_001	13		贝多芬致韦格勒书		75	75	T	0
01140100_001	14		致韦格勒书		77	77	T	0
01140100_001	15		贝多芬致莫舍勒斯书		78	78	T	0
01140100_001	16		思想录			79	T	0
01140100_001	17		关于音乐		81	81	T	0
01140100_001	18		关于批评		85	85	T	0
01140100_001	19		参考书目		89	89	T	0
01140100_001	20		附录			91	T	0
01140100_001	21		贝多芬评传		93	93	T	0
01140100_001	22		贝多芬的作品及其精神		110	110	T	0
01140100_001	23	一、	贝多芬与力		110	110	T	0
01140100_001	24	二、	贝多芬的音乐建树		117	117	T	0
01140100_001	25	三、	重要作品浅释		126	126	T	0
01140100_001	26		论莫扎特		150	150	T	0
01140100_001	27		编者后记		173	173	T	0

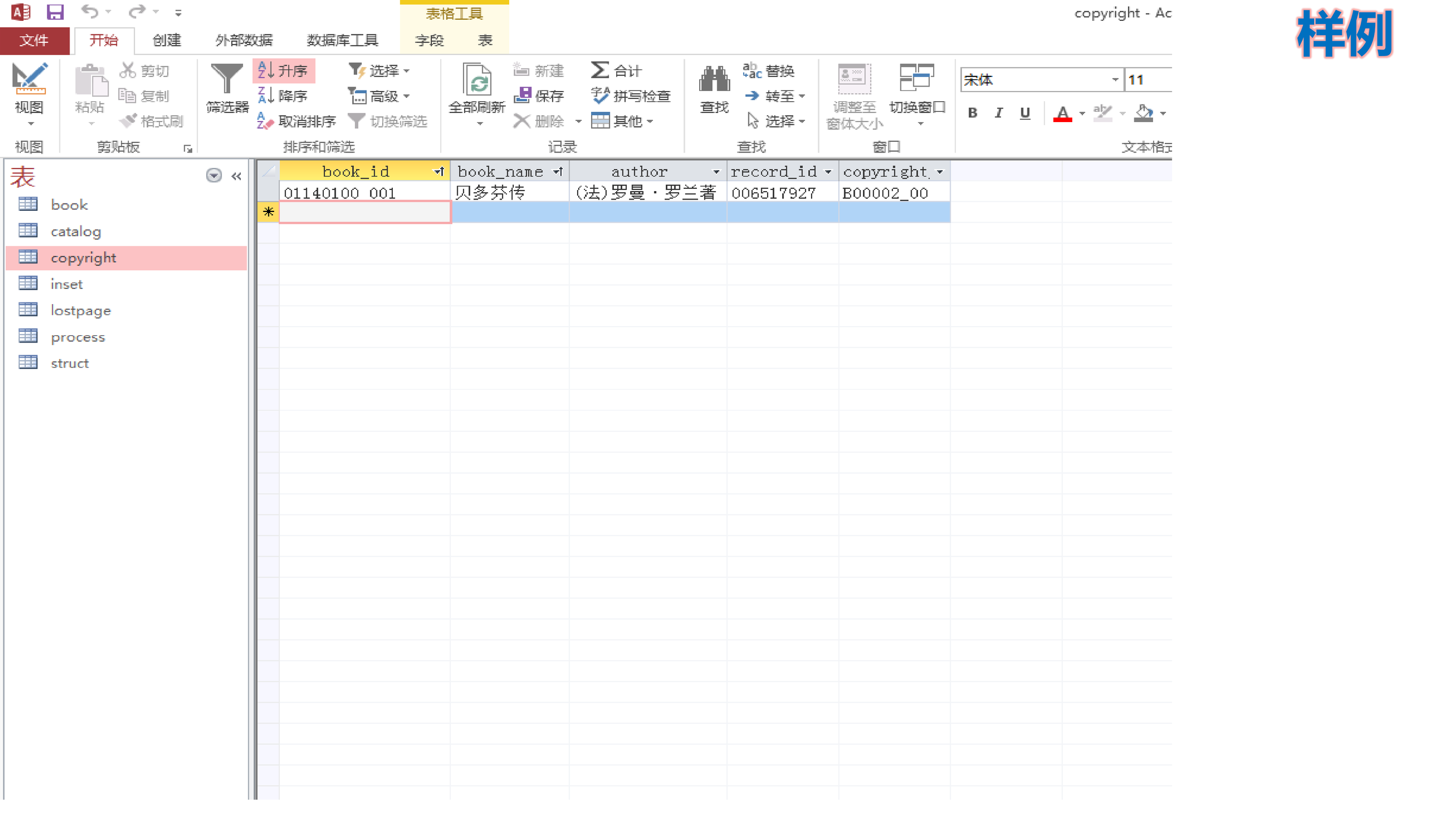
表

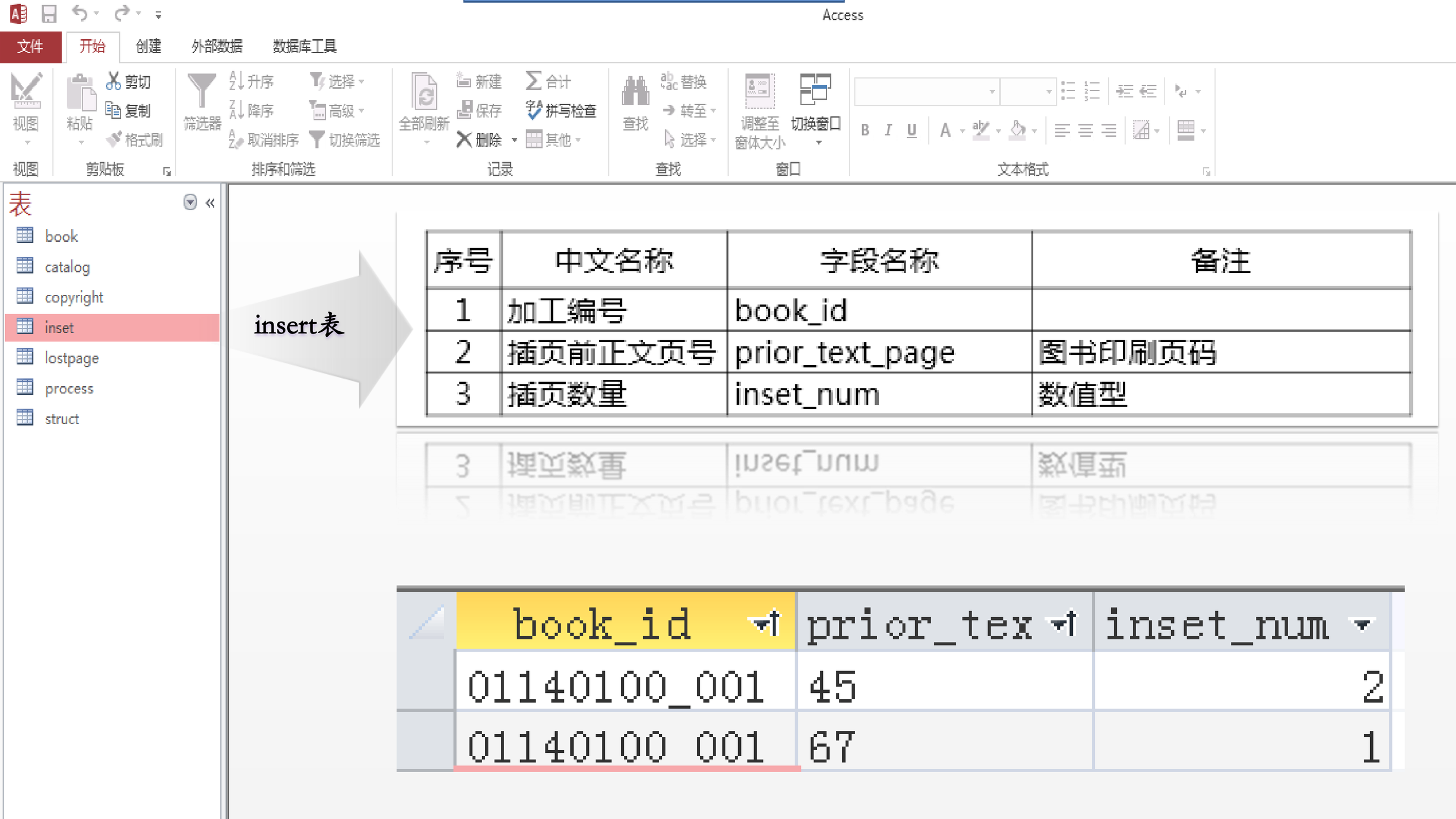
- book
- catalog
- copyright
- inset
- lostpage
- process
- struct

copyright表

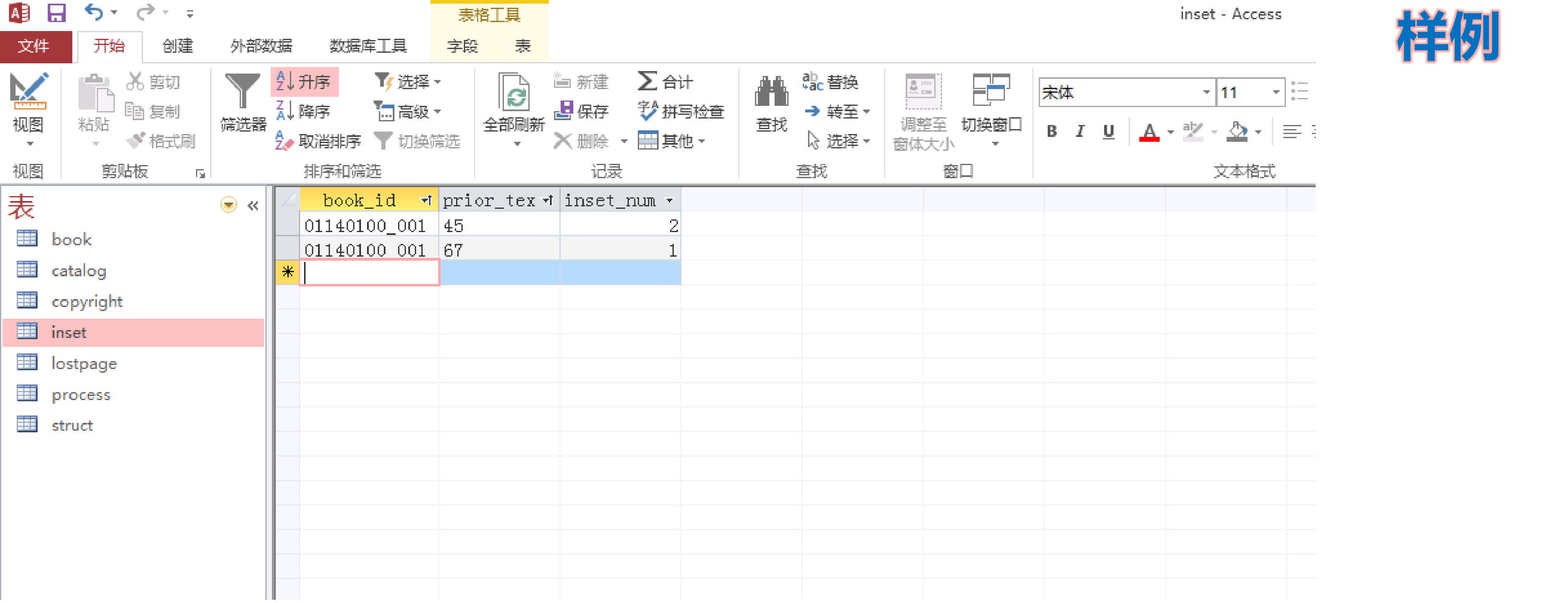
序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	书名	book_name	
3	作者	author	
4	001	record_id	
5	版权页位置	copyright_place	记录版权页文件名

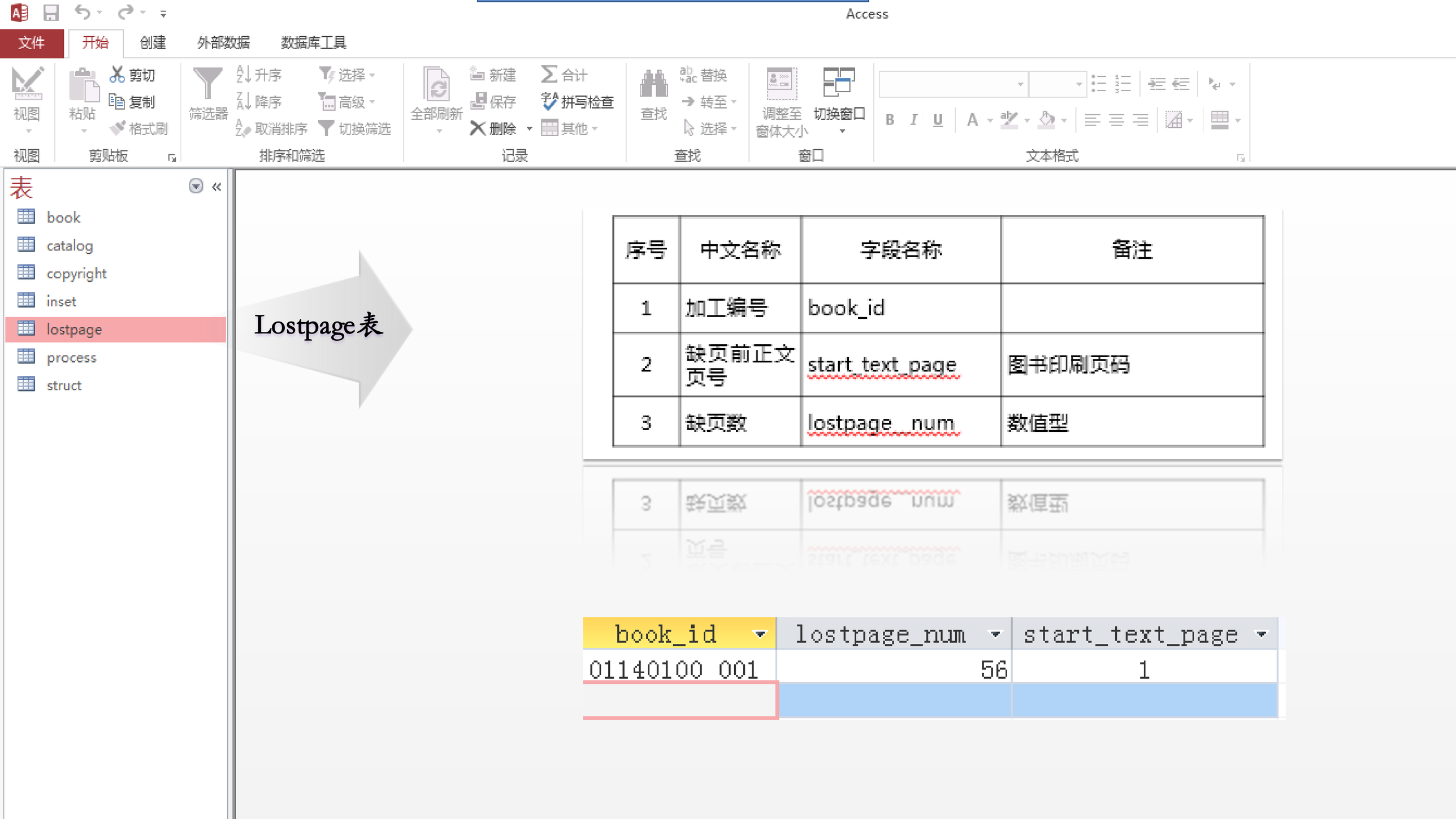
book_id	book_name	author	record_id	copyright
01140100_001	贝多芬传	(法)罗曼·罗兰著	006517927	B00002_00


表格工具
copyright - Ac
样例
文件
开始
创建
外部数据
数据库工具
字段
表
视图
粘贴
剪切
复制
格式刷
筛选器
升序
降序
取消排序
选择
高级
切换筛选
全部刷新
新建
保存
删除
合计
拼写检查
其他
查找
替换
转至
选择
调整至窗体大小
切换窗口
宋体
11
B
I
U
视图
剪贴板
排序和筛选
记录
查找
窗口
文本格式
表
book
catalog
copyright
inset
lostpage
process
struct
book_id
book_name
author
record_id
copyright
01140100 001
贝多芬传
(法)罗曼·罗兰著
006517927
B00002_00


Access
文件
开始
创建
外部数据
数据库工具
视图
粘贴
剪切
复制
格式刷
筛选器
升序
降序
取消排序
选择
高级
切换筛选
全部刷新
新建
保存
删除
合计
拼写检查
其他
查找
替换
转至
选择
调整至窗体大小
切换窗口
视图
剪贴板
排序和筛选
记录
查找
窗口
文本格式
表
book
catalog
copyright
inset
lostpage
process
struct
insert表
序号
中文名称
字段名称
备注
1
加工编号
book_id
2
插页前正文页号
prior_text_page
图书印刷页码
3
插页数量
inset_num
数值型
book_id
prior_tex
inset_num
01140100_001
45
2
01140100_001
67
1

样例

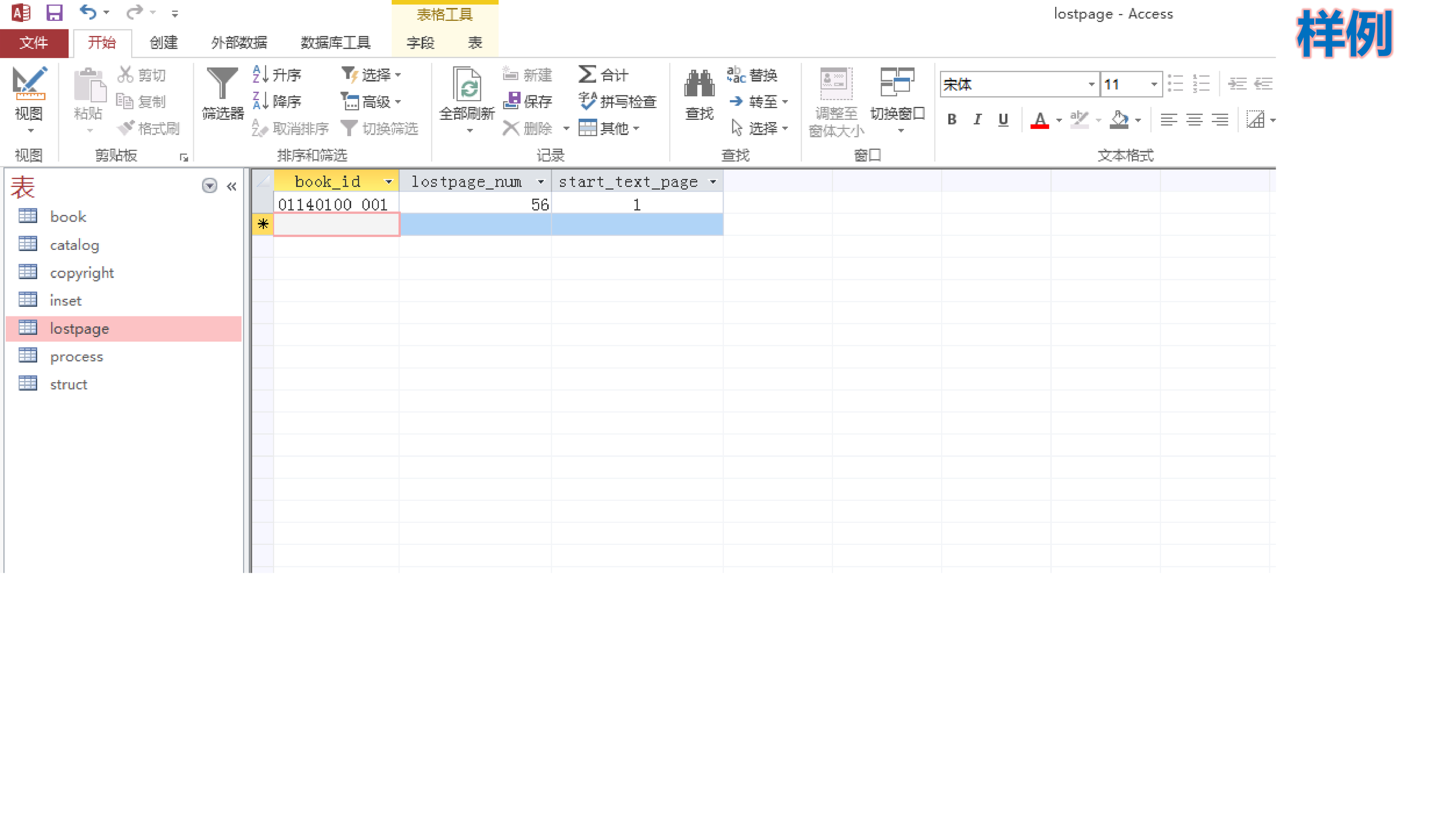

表格工具
inset - Access
文件
开始
创建
外部数据
数据库工具
字段
表
视图
粘贴
剪切
复制
格式刷
筛选器
升序
降序
取消排序
选择
高级
切换筛选
全部刷新
新建
保存
删除
合计
拼写检查
其他
查找
替换
转至
选择
调整至窗体大小
切换窗口
宋体
11
视图
剪贴板
排序和筛选
记录
查找
窗口
文本格式
表
book
catalog
copyright
inset
lostpage
process
struct
book_id
prior_tex
inset_num
01140100_001
45
2
01140100_001
67
1

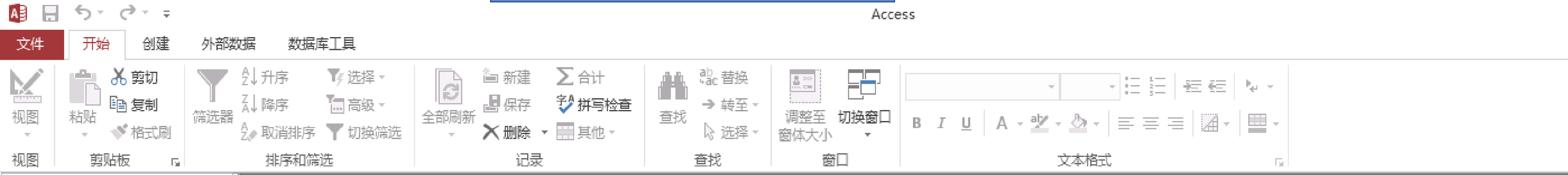


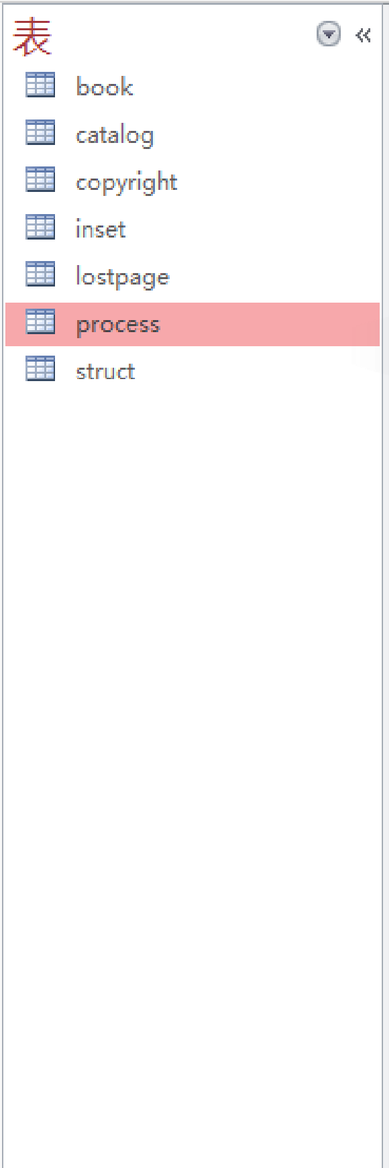
序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	缺页前正文页号	start_text_page	图书印刷页码
3	缺页数	lostpage_num	数值型

book_id	lostpage_num	start_text_page
01140100 001	56	1

样例


表格工具
lostpage - Access
文件
开始
创建
外部数据
数据库工具
字段
表
视图
剪切
复制
格式刷
粘贴
筛选器
升序
降序
取消排序
选择
高级
切换筛选
全部刷新
新建
保存
删除
合计
拼写检查
其他
查找
替换
转至
选择
调整至窗体大小
切换窗口
宋体
11
视图
剪贴板
排序和筛选
记录
查找
窗口
文本格式
表
book
catalog
copyright
inset
lostpage
process
struct
book_id
lostpage_num
start_text_page
01140100_001
56
1





process表

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	书名	book_name	
3	扫描分辨率	dpi	
4	压缩因子	comp_factor	
5	灰度页数量	grey_num	
6	彩色页数量	col_num	
7	TIFF 数量	tiff_num	
8	PDF 数量	pdf_num	包含单版和合并版总数量
9	TXT 数量	txt_num	包含单版和合并版总数量
10	TIFF 存储量	tiff_mb	存储单位：MB
11	PDF 存储量	pdf_mb	存储单位：MB
12	TXT 存储量	txt _kb	存储单位：KB
13	TIFF 硬盘位置	hdA_place	硬盘号
14	PDF 硬盘位置	hdB_place	硬盘号
15	TXT 硬盘位置	hdC_place	硬盘号

book_id	book_name	dpi	comp_facto	grey_num	col_num	tiff_num	pdf_num	txt_num	tiff_mb	pdf_mb	txt_kb	hdA_place	hdB_place	hdC
01140100_001	贝多芬传	300	80	188	4	192	193	193	839	36	152	01140100	01140100	011

样例

process - Access

表格工具

文件 开始 创建 外部数据 数据库工具 字段 表

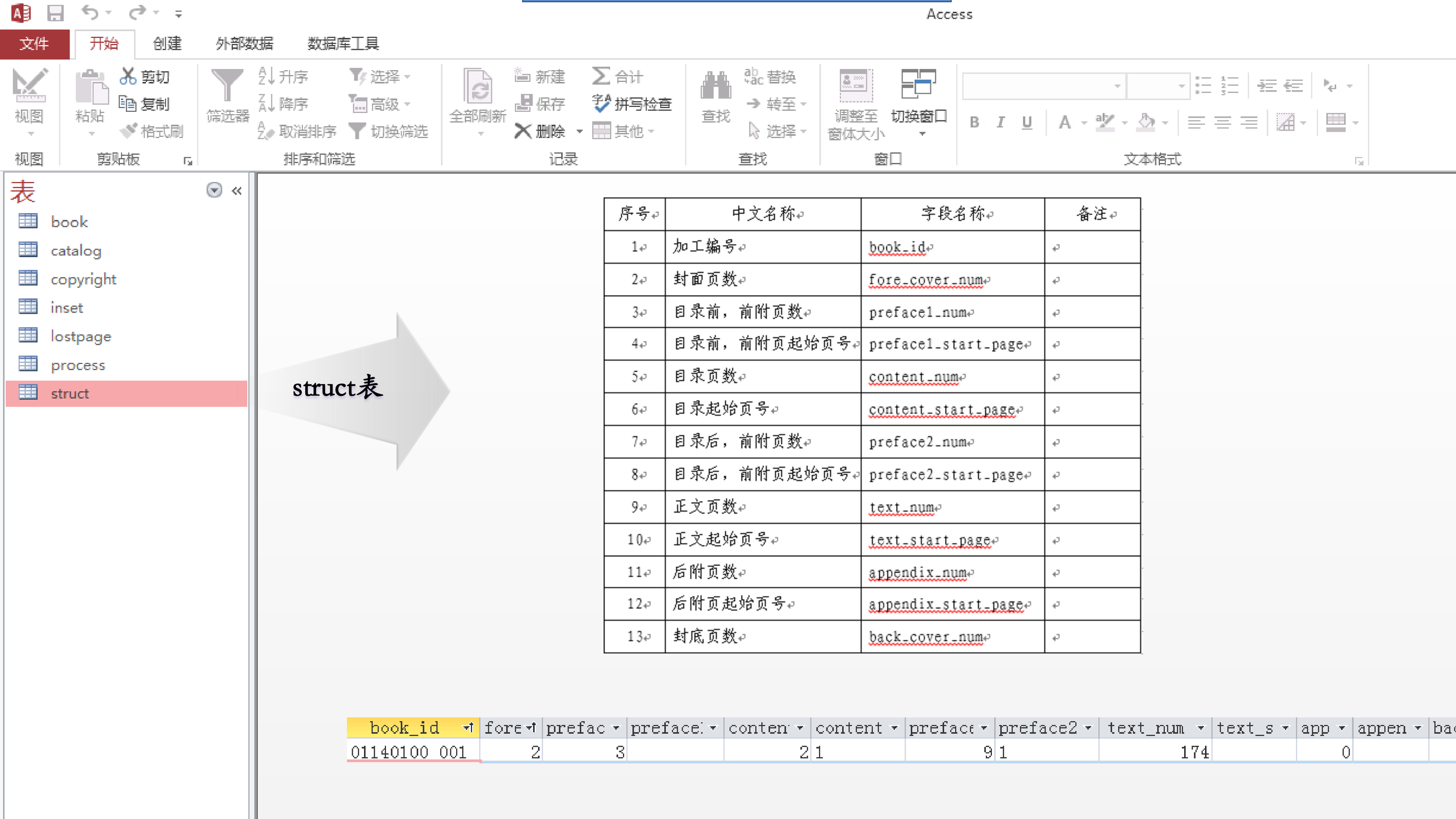
视图 剪切 复制 格式刷 粘贴 筛选器 升序 降序 取消排序 选择 高级 切换筛选 全部刷新 新建 保存 删除 合计 拼写检查 其他 查找 替换 转至 选择 调整至窗体大小 切换窗口 宋体 11

视图 剪贴板 排序和筛选 记录 查找 窗口 文本格式

表

- book
- catalog
- copyright
- inset
- lostpage
- process
- struct

book_id	book_name	dpi	comp	grey	col	tiff_n	pdf_num	txt_num	tiff_	pdf	tx	hdA_place	hdB_place	hdC_place
01140100_001	贝多芬传	300	80	188	4	192	193	193	839	36	152	01140100	01140100	01140100
													0	



序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	封面页数	fore_cover_num	
3	目录前，前附页数	preface1_num	
4	目录前，前附页起始页号	preface1_start_page	
5	目录页数	content_num	
6	目录起始页号	content_start_page	
7	目录后，前附页数	preface2_num	
8	目录后，前附页起始页号	preface2_start_page	
9	正文页数	text_num	
10	正文起始页号	text_start_page	
11	后附页数	appendix_num	
12	后附页起始页号	appendix_start_page	
13	封底页数	back_cover_num	

book_id	fore	prefac	preface	conten	content	prefac	preface2	text_num	text_s	app	appen	ba
01140100_001	2	3		2	1	9	1	174		0		

样例

struct - Access

表格工具

文件 开始 创建 外部数据 数据库工具 字段 表

视图 粘贴 剪切 复制 格式刷 筛选器 升序 降序 取消排序 选择 高级 切换筛选 全部刷新 新建 保存 删除 合计 拼写检查 其他 查找 替换 转至 选择 调整至窗体大小 切换窗口 宋体 11

视图 剪贴板 排序和筛选 记录 查找 窗口 文本格式

表

- book
- catalog
- copyright
- inset
- lostpage
- process
- struct

book_id	fore	prefac	preface	conten	content	preface	preface2	text_num	text_s	app	appen	back_cc
01140100_001	2	3		2	1	9	1	174		0		2

说明文件

文献总体说明文件

数据总体说明
文献单册数据量统计
文献全文转换未转换文件记录表
全文转换加工文字量统计表

单册文献说明文件

图书的名称、作者和页数

存储介质说明文件

存储介质信息：文献数量、文件数量、存储容量等；

技术参数：存储格式、加工设备、加工软件、扫描方式、扫描分辨率等

地方文献数字化项目工作流程

- 文献目录确认
- 数字化加工
- 数据验收

地方文献数字化项目工作流程

数据验收

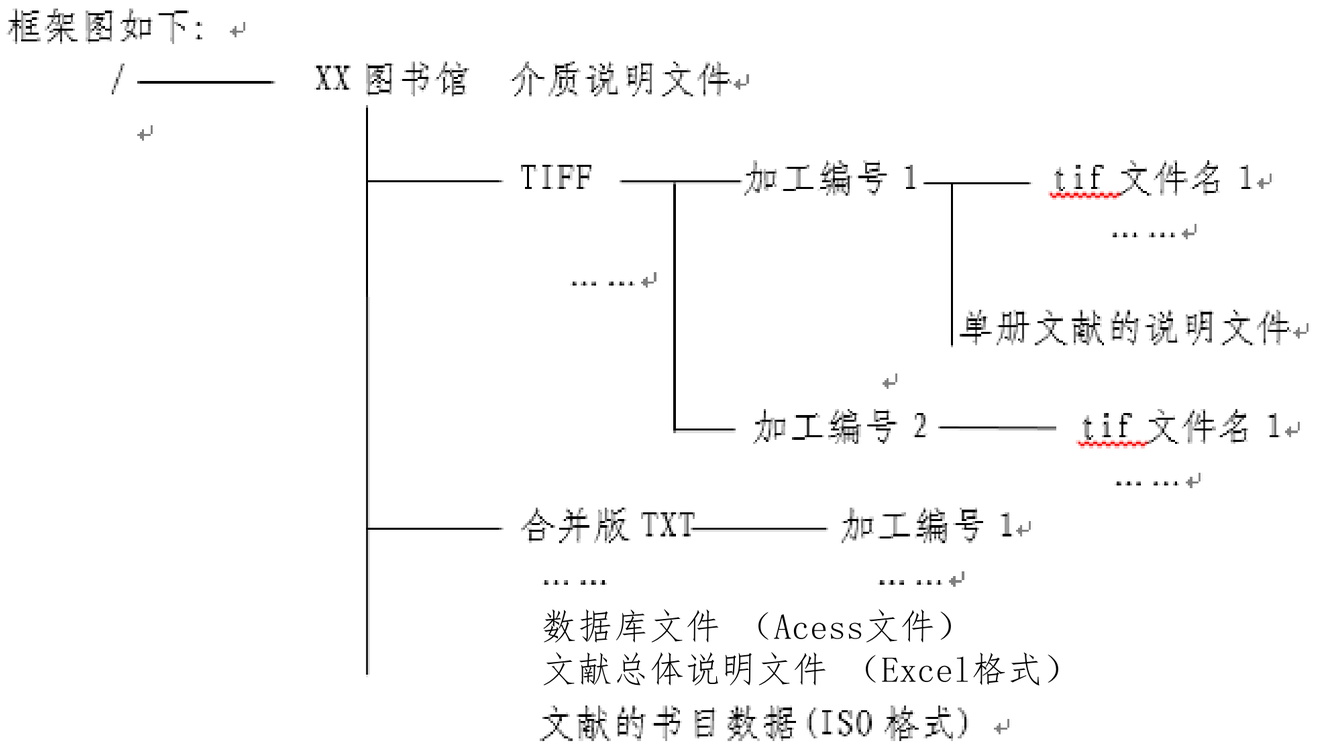
- 规范性
- 一致性
- 正确性
- 完整性
- 有效性

地方文献数字化项目工作流程

数据验收

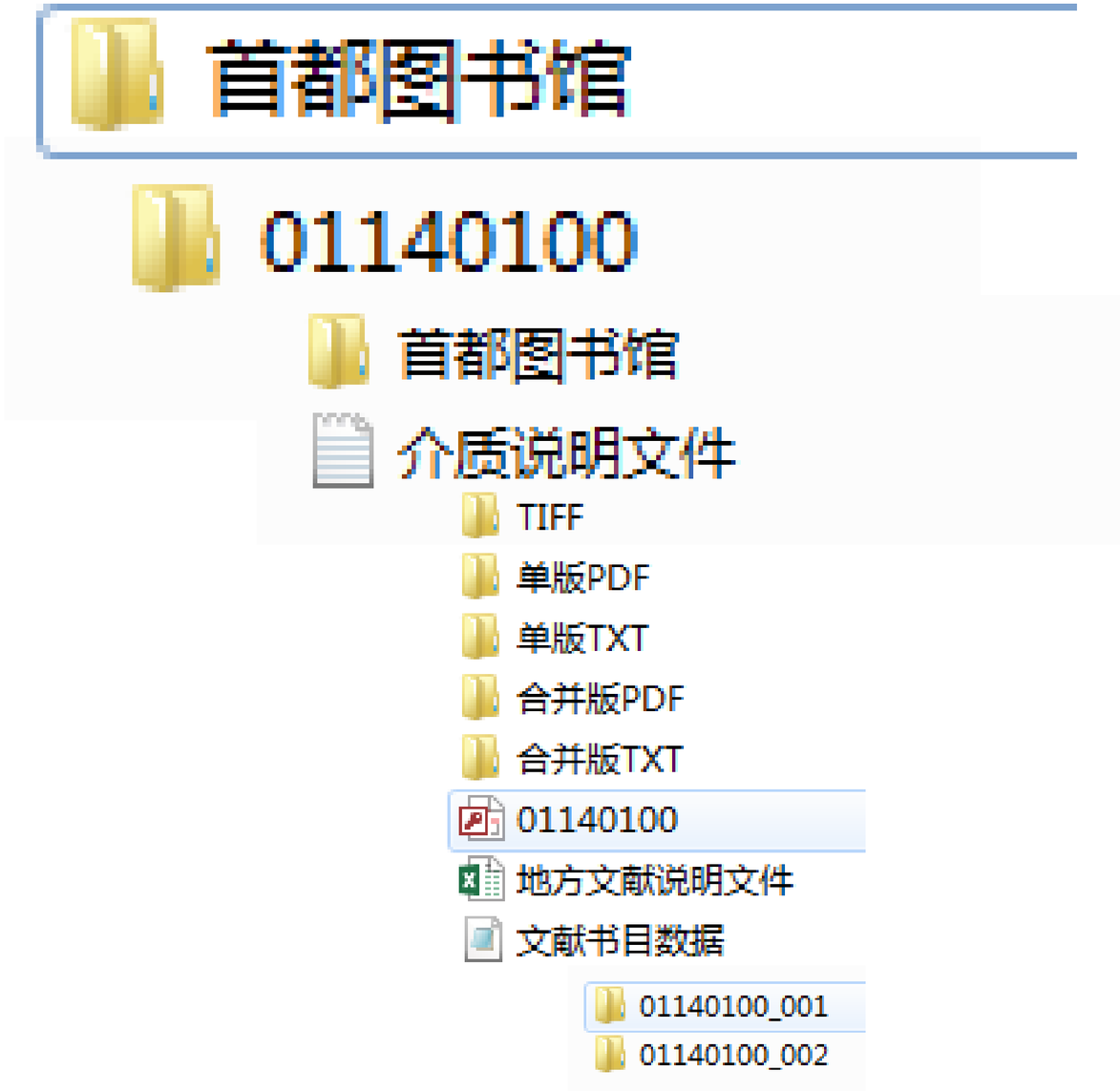
提交格式

框架图如下：



地方文献数字化项目工作流程

首都图书馆为例



地方文献数字化项目工作流程

数据验收

对象数据

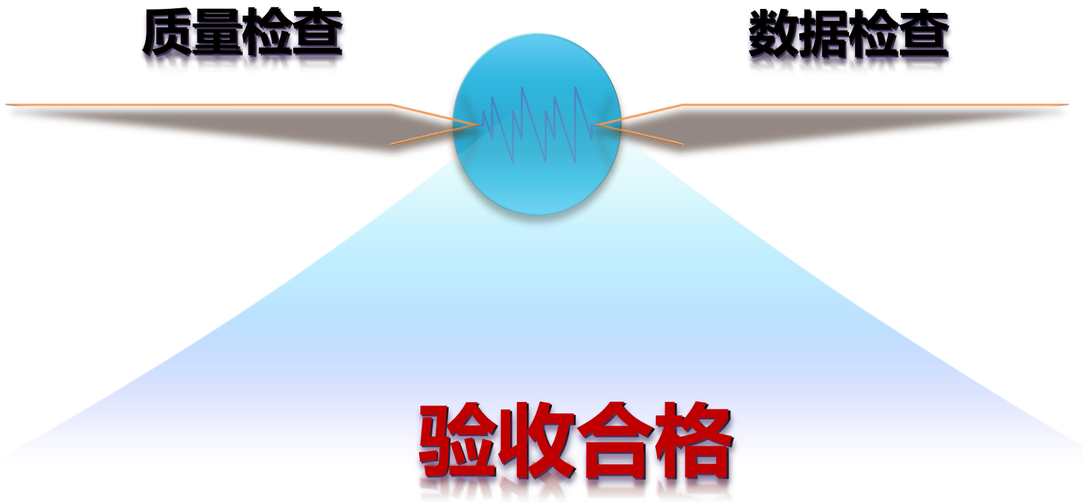
- Tiff图像
- txt文件（单版和合并版）
- 双层PDF（单版和合并版）

数据库及相关说明文件

- 对应数据库
- 介质说明文件
- 地方文献数据说明
- 单册文献说明文件
- 文献书目数据

数据提交相关文件

- 第三方质检报告
- 地方文献验收数据提交单



实例


文献数字化QQ群号
368907080