北京万方数据股份有限公司

询

价

函

二零贰肆年六月

我公司根据目前业务拓展需求，参照政府采购相关规定，就“期刊元数据加工、引文数据加工”进行询价，要求如下：

1. 项目需求（附件一、附件二、附件三、附件四）。
2. 报价单（附件五：请各报价商严格按照报价格式进行报价）。
3. 联系人：刁思路

联系电话： 010-58882757 13051313392

询价单位：北京万方数据股份有限公司

1. 请各单位将制作好的报价单于2024年6月18日下午15：00前送至或快递到我公司。

附件一 中文期刊数据库加工规范

(一)LW表

**1.QCODE**

应与公司提供的期刊标识一致，序号必须按001,002,003……依次入库。

**2.期刊名称(KM)**

与原书期刊名称相对应。

**3.英文期刊名称(KME)**

以公司提供的数据库为准。

**4.年(N)**

与原书一致，封面有手写体的以封面手写体年为准。

**5. 卷(J)**

与原书一致。

**6. 期(Q)**

与原书一致，封面有手写体的以封面手写体期为准。

**7. 页码(Y)**

(1) 库中页码必须与pdf页码相同

(2)格式为：“起始页-末页”。

(3) 当有下转页时，必须规范为：“起始页-末页，下转页”。

(4)当有插页时，必须规范为：“前插XX、中插XX或后插XX”。

(5)由两本或两本以上合为一期的书，须接第一本书往后顺页码，从正文页开始顺。

**8.中英文标题(TM\TME)**

 (1)标题中的破折号“--”必须改为“——”。

(2)标题中不能出现“空格”“\*”、“［1］”、“①”“1)”等无任何意义的字符。

(3)目录中有标题而正文中没有标题时,按目录标题入库。

（4）中英文标题均提取。

9.中英文引题（YT\YTE）

（1）引题中不能出现“空格”“\*”、“［1］”、“①”“1)”等无任何意义的字符。

（2）引题中的破折号“--”必须改为“——”。

（3）中英文引题均提取。

10.中英文副题（FT\FTE）

(1)副题中的破折号“--”必须改为“——”。

（2）副题中不能出现“空格”“\*”、“［1］”、“①”“1)”等无任何意义的字符。

（3）中英文副题均提取。

**11.中英文摘要(ZY\ZYE)**

(1)中文摘要字段开头不能出现字段名，例如[摘要]、摘要：、摘要等。

 (2)英文摘要字段开头不能出现字段名，例如[Abstract]、Abstract：、Abstract等。

(3)字段开头不能出现空格

(4)摘要中出现“目的……”、 “方法……”、“ 结果……”、“ 结论……”时，如果这几个词后有标点符号，按原书即可，如没有，需在这几个词后分别添加半角空格。

（5）中、英文摘要的段落格式参照原书。

（6）中英文摘要均提取。

**12.中英文关键词(GJ\GJE)**

(1)中文关键词字段开头不能出现字段名，例如[关键词]、关键词：、关键词等。

 (2)英文关键词字段开头不能出现字段名，例如[Key World]、Key World：、Key World等。

(3)字段开头不能出现空格

(4)若某一篇分隔符全部为“逗号”、“顿号”、“空格”时，此篇全部改为“分号”。

(5)若某一篇分隔符中有“分号”和其他标点符号时，此篇分隔符与原书一致。

(6)关键词末尾不能有标点符号。

(7)中文关键词除了含有特殊意义的，字与字之间不可以存在空格。

（8）中英文关键词均提取。

**13.中图分类号(FL)**

(1)格式为:大写字母+数字，分类号必须以大写字母开始，大写字母与数字间不允许出现空格。

(2) 只有数字或数字开头的也作为中图分类号，按原书提取。

(3)只有一个大写字母时也做为中图分类号。

(4)两个中图分类号之间用“分号”隔开。

（5）中图分类号中不能有空格。

**14.文献标识码(BS)**

文献标识码为大写字母。

**15.文章编号(BH)**

文章编号中不可存在空格。

**16.基金项目(JJxm)**

（1）基金项目一般由项目名称和项目编号组成。

（2）凡含有基金项目特征词的内容均提取。

（3）若“基金项目”和“致谢”中的项目名称、项目编号内容一样时，“致谢”不用保留，不一样是都需要保留。

（4）中英文基金项目均提取。

**17.收稿日期(SG)**

（1）格式为：XXXX年XX月XX日

（2）收稿日期提取规则：收文日期、收稿日期、收到日期、到稿日期、收到初稿日期、接收日期、收入日期、Date of reception(收稿)、Manuscript received（稿件接收）、Received\_Date（收到）等。

（3）有年月无日，日补录1；只有年（无月和日）、只有年和日、不符合日期逻辑的均不提取，例如：“2023-10”规范为“2023-10-1”，“2023- -5”不提取，“2023”不提取，“2023-14-5”不提取。

**18.修稿日期(XG)**

（1）格式为：XXXX年XX月XX日

（2）修稿日期提取规则：修定日期、修复日期、修订日期、改回日期、改稿日期、收修订稿日期、修改稿返回日期、修改稿收到日期、收到修改稿日期 、修改稿日期、接受日期、录用日期、收修改稿日期、修订回稿日期、返修日期、修回日期 、Accepted（接受） 、Accept\_Date（修改）、received revised（收到修订）、revised（修订）等。

（3）有年月无日，日补录1；只有年（无月和日）、只有年和日、不符合日期逻辑的均不提取，例如：“2023-10”规范为“2023-10-1”，“2023- -5”不提取，“2023”不提取，“2023-14-5”不提取。

**19.中英文栏目(LM\LME)**

(1)按目录入库。

 (2)全书目录中无栏目时，不用加栏目。

(3)如果大栏目与小栏目同时存在时，全书按小栏目入库。

 (4)不同大栏目下小栏目存在重复时，按大栏目入库。

 (5)合刊时，栏目重复，栏目应分别规范为“栏目1”和“栏目2”。

（6）同一大栏目下的小栏目重复，加小栏目。

**20.正文(ZW)**

(1)如文章无摘要时留取100字左右正文，以句号或**明显结束符**结尾。

(2)正文从第一段开始留起。

(3) 正文段落格式按原书提取。

**21.文章序号(DOI)**

按原书即可。

**22.pdfurl**

 年后两位+期+01(序号).pdf

**(二)YW表**

1.参考文献内容（TM）

（1）参考文献和注释全部著录在参考文献字段。

（2）凡属本篇论文的参考文献和注释内容全部提取。

（3）所有注释、参考书目、参考资料、引文文献、参考 数据等内容都作为参考文献。

（4）若一条参考文献中，不能识别的小语种占一半以上，则该条参考文献内容不识别，只需要著录该条文献的“序号+不可识别小语种”，例如：“[2]不可识别小语种”。若一条参考文献中，有个别不能识别的小语种文字，则该部分用括号“（）”表示，其余能识别内容正常录入。

（5）参考文献和注释的序号格式与原书保持一致，若注释序号超过㊿无法录入的，序号格式改成与参考文献格式不同的格式，优先改成“（50）”的格式，若参考文献序号格式为小括号的，注释格式改成“50.”的格式。

（6）若参考文献开头出现“——”时，代表该条参考文献的人名与上一条一致，须复制上一条的人名替换该条的“——”。

（7）参考文献/注释内容提取顺序按原书录入。

2.参考文献序号（CX）

每一篇参考文献条数顺序号从1开始，依次排序。

3.QCODE

（1）应与公司提供的期刊标识一致，序号必须按001,002,003……依次入库。

（2）与LW、ZZ对应该篇论文标识号与一致。

(三)ZZ表

**1.中英文作者(ＺＺ)**

(1)作者后不能出现“\*”、“△”、“+”、“，”、“数字”等字符。

(2)当人名前出现“本刊记者”、“本刊”、“通讯作者”、“本刊评论员”、“文/”等字样时，需将此字样删除。

(3)中文作者中不能出现空格，除含有特殊意义除外。

(4) 下列词后的人名是作者：文、译、翻译、著、整理、口述、采编、撰稿、执行、策划、综述、本刊评论员、本刊记者、设计、绘画、编绘、审校、指导、图作者、摄影作者、主持人、据XX、来源XX、摘自XX、笔名、佚名等。

(5)作者中不允许出现职称。

（6）标题下有正规作者时，文章末尾的执笔、供稿、整理、等不提取为作者（撰文、译等除外，需要提取为作者）。

（7）标题下无正规作者时，文章末尾的执笔、供稿、整理、撰文、译等提取为作者。

（8）作者为集体作者（课题组）的，提取集体作者（课题组）名称、含有作者简介的成员姓名、译、撰文等为作者，执笔、整理等不提取为作者；“本刊编辑部”当作集体作者时,需规范为《本刊名》编辑部。

（9）中英文作者应对应；中英文作者顺序不对应时，英文作者应按中文作者顺序一一对应；当中英文作者个数不对应时，中文作者和英文作者全部提取，（即有几个要几个）。

（10）中英文作者均提取。

**2.中英文作者单位(DWys)**

(1)一般格式为：“实体单位”+“城市”+“邮编”。

(2) 实体单位与省市、邮编间必须有分隔符，原书有按原书，原书没有统一用“逗号”或“空格”隔开；省市和邮编间不作此要求，按原书。

(3)中文作者单位末尾的标点符号需删除。

(4)单位中不能出现序号。例如1.2.3.或1 2 3。

(5)作者单位中不可出现职称或学历。

(6)单位中多余内容必须删除。

(7)单位不完整时需要补全。例如“技术中心”“本刊编辑部”“XX报”等

(8)如单位为“河北省省长” ，提取并规范为“河北省政府”

(9) “通讯地址”、“联系地址”都不当单位，除英文单位和中文通讯地址互译外

a.XX级不为单位。

例：“清华大学2009级，北京，100088”应规范为：“清华大学,北京，100088”。

b.XX专业不为单位。

例：“广州民航职业技术学院飞机结构修理专业”应规范为“广州民航职业技术学院”。

(10)两个及以上作者单位，用分号隔开，分号前后不能出现空格。

（11）中文论文的中文单位按作者的对应标记进行匹配，无明确的对应关系时，首先参照作者简介匹配单位；若不能参照作者简介的，作者与单位个数相同的，须依次为每个作者匹配一个单位；作者与单位个数不同的，无法区分的，须将所有单位匹配给每个作者。

（12）中文论文提取英文作者单位，英文单位与英文作者若有非常明确的对应关系的，则精确匹配；若对应关系模糊的（需人工判断），则将所有的英文单位匹配给每个英文作者（模糊匹配的英文单位，入库时在每条单位前加“@@@”）。

（13）英文论文的英文单位按作者的对应标记进行匹配，无明确的对应关系时，首先参照作者简介匹配单位；若不能参照作者简介的，作者与单位个数相同的，须依次为每个作者匹配一个单位；作者与单位个数不同的，无法区分的，须将所有单位匹配给每个作者。

（14）标题下正规位置没有单位，从作者简介、作者地址、作者通讯地址中提取所有的现单位，原单位不提取。

（15）中英文作者单位均提取。

**3.中英文作者简介(ZJ)**

(1) 中英文作者简介要与中英文作者一一对应。

（2）凡属作者信息的内容，均提取为作者简介。（除只有单位或省市、邮编的情况）

（3）正文中无简介，也不从人物介绍中提取简介。

（4）若作者简介内容只有身份证号时，不提取为作者简介。

（5）中英文简介中必须有对应作者姓名。

（6）中英文作者简介均提取。

4.作者角色（JS）

（1）作者角色字段，标记4种角色：“通讯作者、集体作者、第一作者（作者简介中明确提到的）、译”。

5.QCODE：

（1）应与公司提供的期刊标识一致，序号必须按001,002,003……依次入库。

（2）与LW、YW对应该篇论文标识号与一致。

6.作者序号（CX）:

（1）每一篇中英文作者顺序号各从1开始，依次排序。

7.中英作者标记（CH）：

（1）中文作者画√，英文作者不画√。

附件二 外文期刊加工标准

**一、标题:**

（一）标题以原书内容为准。

（二）标题末尾出现的符号只保留缩写点、问号、感叹号、省略号，其他符号删除；标题中双引号需改为单引号。

（三）标题中出现前后带“—”或 “～”的副标题时，需将前一个“—”或 “～”改成冒号加半角空格，后一个“—”或 “～”删除。

（四）标题中出现 ™ ,®,6th, 3rd, 1st, 2nd不用加上标

**二、作者**

 作者之间用“；”分割，没有作者不用加分号

仅有作者时，地址、E-mail必须有对应分号，三者为对应关系

（一）作者内容以原书内容为准。

（二）作者中不允许出现职称。

（三）作者中不允许出现逗号，如有应改为空格。

（四）作者中双引号需改为单引号。

（五）超过40个作者（含40），只取前5个，其余作者不要，用一个“et al”代替。

（六）录大标题时，只提大标题下的作者。

（七）录小标题时，优先取小标题下的作者，小标题下无作者时取大标题下的作者。

（八）团体性的作者不取，如XX委员会、XX组等。

**三、地址**

（一）地址内容以原书内容为准，专用名词除外如“eBBB”。

（二）地址末尾出现的符号只保留缩写点，其他符号删除。

（三）地址中双引号需改成单引号。

（四）无正规地址时，可从作者简介中查询有无可提取的地址。

（五）地址中出现“Present address:”代表是现在的地址。

（六）同一作者出现两个地址或单位的，两个单位之间用”##”隔开。

**四、E-mail**

（一）取E-mail时，应从“E-mail:”的后边开始内容取，如“E-mail: daniele.gatti@unipv.it”只取daniele.gatti@unipv.it即可。

（二）无法判断E-mail归属为哪个作者时，只给第一作者。

（三）有作者无E-mail时，E-mail必须用分号代替。

（四）一个作者有两个E-mail时，两个E-mail之间用空格隔开。

（五）E-mail中的标点符号只允许有“.”。

（六）E-mail内容结尾不允许有标点符号。

**五、摘要：**

（一）摘要内容以原书内容为准。取摘要时优先取正规摘要“Abstract”、“Summary”下的内容，无正规摘要时按“Conclusion”，无以上三种情况时取正文第一段，选取内容到整段结束，夹在文中的小标题不要。

（二）正文中没有正规摘要，书中有单独的摘要页，应从摘要页选取相对应的摘要。

（三）摘要中出现的网址或E-mail应删除连接的整句话，若网址或E-mail出现在括号中，则删除括号及里面的内容。

（四）摘要末尾无标点符号时应加“.”。

 （五）摘要中代表注释、参考文献的数字角标应删除。

**六、关键词：**

（一）关键词和关键词之间用“分号+半角空格”隔开。

（二）关键词最后所有的符号都不要，但特殊的缩写点保留，如国家地区缩写等。

（三）关键词大小写与原书一致。

（四）关键词前表示排序的数字或字母应删除。

（五）关键词全部为代码时，查OCIS表或PACS表，查不到的则把该代码删除

**七、页码：**

（一）文章的页码必须包含“起始页”和“结束页”。

（二）文章页码是连续的，由“起始页-结束页”组成。

（三）文章页码是不连续的，由“起始页，中间页，结束页”组成。

（四）文章页码只有一页时，也应按照“起始页-结束页”组成。

（五）如有单独摘要页且正文中无正规摘要，填写页码时应加上相对应页码。

（六）正文页码是罗马字母的，。

（七）空白页、广告页不要。

（八）若在目录上提取第二语种标题时，文章页码不用填写目录页页码。

八**、Lan（语种）**

根据正文的内容判断语种

（一）文章内容是两种语言互译（必须正文互译）的，按两条入库。

（二）文章内容有摘要互译，按一条入库。正文语种为第一语种，互译摘要为第二语种，第二语种内容只取标题、摘要、关键词，第二语种语言需在库中填写。

（三）文章内容没有摘要互译，按一条入库。第二语种内容只取关键词，第二语种语言无需填写。

（四）当文章有多种语言时，只选取两种语言，顺序为：英文，德文，法文，其它。

（五）正文中第二语种无标题时，需从目录上提取第二语种的标题。

**九、DOI**

（一）取DOI时，应从“DOI:”后边开始取。

（二）DOI中不能出现空格。

（三）网址型DOI，应从10.XXXX 开始取。

**十、参考文献**

（一）每条参考文献开头加@@。

（二）参考文献中出现的网址、DOI必须回行且不能有空格。

（三）参考文献按第一语种加工。

**十一、其他要求**

1.第二语种

取第二语种的要求

只要标题、摘要、关键词

如只有标题、关键词。只要关键词。不取第二语种（lan\_ab）

2.PDF

Pdf 应于库对应。

PDF做成两份，一份是整本PDF的，一份是按文章分篇的PDF。

3识别规则：

* 一定要保证标题、作者、地址、E-mail这几项的正确率并保证作者、地址、E-mail、这几项的对应关系
* E-mail保证与作者相符合（对应关系）
* 标题、摘要中的特殊字符如α β γ，识别为a,b,y,应改为α β γ，做上下标超过2个字符的用括号括起。确保地址中不能存在职称等不是地址的内容，摘要要以句点结束为准
* 做上下标的标准为先下后上
* 关键词中大小写更改（除国家名、化学符号其它都要是小写）词与词之间要用“;”号隔开末尾的句点不要。

4.**上下角标的转换规则**

上标用~表示

下标用\_表示

**注意：**

原文中ä要改成ae（大写改大写，小写改小写）, ò õ 改原字母，ß要改ss,出现这个α、β字符保持不变

**加工说明：**

为规范外文杂志类、信息类期刊入库标准，便于入库加工过程中文章篇目的选取，此类期刊取入库篇目时，以原书目录为准，目录上有的篇目都需要入库，以下特殊情况需要注意：

1、目录中广告、勘误表、作者简介、下期预告、索引类内容不入库。（作者简介只是不单独入库，但包含的作者E-mail、单位等内容需要入库，）

2、目录中书讯、书评内容走小条。

3、目录中新闻、信息类内容统一走栏目，不单独入库，如个别文章有作者信息（包括作者、E-mail、单位），按正文顺序将作者信息入库。

3.1 原书目录中没有以新闻或信息为栏目，但正文中有的，按栏目入库。

3.2 信息类内容主要指会讯、产品介绍等有关内容。

3.3在新闻栏目下，同一作者有多篇文章的，入库时保留一个即可。

3.4不是所有带“NEWS”单词的就是新闻，需要根据文章内容判断，如出现“NEWS & VIEWS”应该按正规文章入库，而“NEWS IN BRIEF”就要新闻类文章入库。

4、目录中没有但在正文中出现的文章，有作者的正常入库，没有作者的不单独入库，将该篇内容放在邻近篇目一起带过。

5、目录中只有大栏目，但正文中为小篇文章的：

5.1 如果所有小篇文章有作者，每篇文章单独入库；如果有的有作者，有的没有，将没有作者的放在邻近有作者的篇目后一起带过；

5.2 如果所有小篇文章都没有作者，走大栏目；

5.3 目录中有专题、报告类栏目，在栏目正文中有前言、导读之类内容，没有标题且内容超过1页的，以“该专题或报告名+Preface”为标题入库。

6.原书没有目录的，正文内容分为小篇且有作者的，按原文页码顺序入库。没有作者的，整本走一条，以刊名作标题。

7．整本书只有一个作者，目录标题分级或原书没有目录，整本走一条。

以上内容主要针对外文中杂志、信息类期刊目录不规范，加工过程中选取篇目容易出现歧义等问题制定。如今后遇到新的问题，需经各工序协调解决，并制定相应合理的规则对本规则进行补充。

附件三 图像文件处理内容、规则及质量要求

扫描的图像文件按条码命名的文件夹分目录分别存放，文件夹名称与图像中的条码一致。

1、 检查扫描的图像文件，文件夹名称应与第一个图像中的条码相一致，对于缺页情况进行记录，记录内容为：文件夹名称、错误说明。每个文件夹一个说明文件，文件名为：文件夹名称.TXT。
2、根据页面内容选择正确的图像模式和扫描的图像文件，线图优先。彩色图像采用jpg格式，线图、灰度tif格式。
3、成品图像非文字图片区不得有黑边，字里行间黑点不得超过字体1/4大小。
4、成品图像倾斜小于0.5度。
5、成品图像上下左右居中。
6、需通过图像参数调整，使页面文字及图像清晰度最佳，不得出现网底。
7、所有图像分辨率均以原扫描数据的分辩率一致，不得修改。
8、文件大小要求：线图800KB-1MB、灰图为8MB左右、彩图为300-500KB。
9、 成品图像文件命名及存放规则：
成品图像文件按文件夹分目录存放，文件夹命名同原始文件夹名称。图像文件命名顺序同印刷顺序，以6位数字开头，页码不足六位前端以零填充。中文目录、英文目录、封面、没有页码的插页图像文件名中需增加特殊标识，即，在6位数字后，加 \_fm01表示封面，\_fm02 表示封二、\_ml01表示中文目录第一页、\_ml02表示中文目录第二页……，\_mle01表示英文目录第一页、\_mle01表示英文目录第二页……、对于在正文中间未占页码的插页图像文件，文件名的前六位同上一个有页码的图像文件名的前六位，并加\_CY01以标识……，正文结束后未占页码的图像文件（无论是否为广告页）文件名的前六位同上一个有页码的图像文件名的前六位，再加\_01、\_02……以标识。
10、数据按批加工，整批发出，整批返回，每批图像数量在1万-3万之间，加工时限为3个工作日。
11、依批次，按页计，错误率控制在千分之二以内。

附件四 其他加工要求

1.OCR识别错误率

元数据部分：要求文字识别错误率低于万分之五

正文部分：要求文字识别错误率低于万分之五

2元数据字段错误率

元数据字段错误率低于千分之一

|  |  |
| --- | --- |
|  附件五 **报价单** |  |
| 供应商： |  |  |  | 日期 |  |
| 公司地址：  |  |  |  |
| 省/市邮编：  |  |  |  |  |  |
| 电话： 传真： |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **特为下面客户报价** |  |  |   |
|  |  |  |  |  |  |
| 单位名称： | 北京万方数据股份有限公司 |
| 单位地址： | 北京市海淀区复兴路15号 |
| 省/市邮编 | 100038 |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |  |
| **特别注意事项：** |  | 参看附件 |  |  |  |
|

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **资源类型** | **单位** | **报价** | **价格是否可议** | **税率** |
| 中文期刊元数据 | 元/条 |  |  |  |
| 中文期刊引文数据 | 元/条 |  |
| 中外文期刊引文拆分 | 元/条 |  |
| 中外文期刊基金拆分 | 元/4条 |  |
| 中外文期刊作者单位拆分 | 元/4条 |  |
| 中外文期刊作者简介拆分 | 元/4条 |  |
| 英文期刊元数据 | 元/条 |  |
| 英文期刊引文数据 | 元/条 |  |
| 图像修图 | 元/页 |  |
| 日文期刊元数据 | 元/条 |  |
| 俄文期刊元数据 | 元/条 |  |

报价文档提供**3套**如您有任何疑问，请即联络：刁思路 电话：010-58882757 13051313392电子邮件：dsl@wanfangdata.com.cn |

年 月 日

 供应商名称：（加盖公章）