

技术服务有限公司

(2017-2019年)

项目编号:

2 方：知识产权出版社有限责任公司

鉴定地点：北京市

簽定日期：

目 录

第一条 术语和定义.....	2
第二条 技术服务内容.....	3
第三条 技术服务标准.....	5
第四条 技术服务质量.....	5
第五条 合同价.....	6
第六条 考核和验收.....	7
第七条 乙方合同义务转让禁止以及乙方保证.....	8
第八条 保密约定.....	8
第九条 知识产权.....	9
第十条 风险责任.....	9
第十一条 违约责任.....	9
第十二条 合同变更或解除.....	10
第十三条 争议解决.....	11
第十四条 合同附件.....	12
第十五条 合同生效.....	12
第十六条 联系方式.....	12
第十七条 其他约定.....	12
附件.....	1

国家知识产权局专利局(以下简称“甲方”)与知识产权出版社有限责任公司(以下简称“乙方”)根据《中华人民共和国合同法》等相关法律法规的规定,就“中国专利数据标准化加工项目”的技术服务工作,同意按下述条款和条件签署本项目技术服务合同(以下简称“合同”)。

第一条 术语和定义

除非根据上下文另有解释,本合同中使用的术语含义如下:

1、“合同”系指:本合同正文全部条款及其所有的附件和上述文件所提到的构成合同的所有文件。

2、“合同价”系指:根据本合同规定乙方在正确地完全履行合同义务后甲方应付给乙方的价格。

3、“服务”系指:根据本合同由乙方履行的任务和服务,其内容包括但不限于:数据接收、数据分析、加工流程设计、数据处理、数据质检、数据纠错、数据提交、技术服务工作报告、验收、咨询、技术支持和项目管理等活动。

4、“合同产品”系指:根据本合同规定的全部可交付物,其包括乙方应交付给甲方的全部数据产品、文件、文档和其它材料。

5、“数据资源”系指:在本合同履行过程中使用的所有原始数据、成品数据或半成品数据。

6、“数据产品”系指:乙方履行本合同提交给甲方的成品数据。

7、“XML”系指:可扩展标记语言。

8、“验收”系指:甲方或甲方指定的第三方根据本合同所规定的验收程序和条件,对数据产品进行检验和测试,以验证其是否满足本合同所规定的标准及合同附件的要求。

9、“资料”系指:根据本合同规定,由乙方提供的文件清单、文件、报告、图表等文字资料。

本合同及附件中所涉及的有关术语和定义,以国家有关标准为准。没有可参照的,双方自行约定并作为附件。

第二条 技术服务内容

1、数据源

2016年11月1日至2019年10月31日公开的中国发明公布、发明授权和实用新型授权的单行本原始数据。预计加工量约为765万到900万件，最终加工量以实际发生为准。乙方应对源数据进行校验，验证源数据的完整性、可用性和一致性。乙方对于发现的源数据错误应作标记，并保证及时通知甲方。

2、技术服务时间

乙方应在2019年11月10日前完成上述数据源的全部加工工作并提交数据产品；在此期间内，乙方须按照本合同约定和要求，为甲方提供技术服务并向甲方交付合同产品。

3、技术服务实施路线

3.1 加工内容

3.1.1 标准化文本数据加工

3.1.1.1 获取完整数据源

通过FTP或网络共享或移动硬盘等方式获取中国专利文献单行本原始数据，通过数据解压、校验及解析等手段验证原始数据包的完整性、可用性和与公报数据的一致性。

3.1.1.2 格式标准化处理和整合

1) 按照《专利数据元素标准第2部分：关于用XML处理中国发明、实用新型专利文献数据的暂行办法》(ZC 0012.2—2006)的要求，逐件逐页分别对著录项目、名称、摘要、摘要附图、权利要求书、说明书、说明书附图进行数据格式标准化处理，形成符合ZC 0012.2—2006标准的著录项目和说明书全文数据。

2) 对中国专利文献数据(包括各种内容、各种格式等)进行整合，规范存储方式、存储路径，建立数据增加、修改、连接、提取等功能，形成一套完整、准确的、符合ZC 0012.2—2006相关规定的中国专利基础数据。

3.1.1.3 内容标准化处理和整合

1) 按照《专利文献数据规范》(ZC 0014—2012)，根据甲方的业务要求，对完成格式标准化处理和整合的中国专利基础数据进行以专利为单位的拆分和校验。

2) 对拆分后的数据同《专利文献数据规范》(ZC 0014—2012)的结构进行对应和转换，形成以专利为单位的XML文件，同时对XML的内容进行必要的规范和转换，从转换后的XML文件中提取数据内容，保存到关系数据库中。

3) 按照甲方的业务要求,利用程序实现从关系数据库中提取著录项目字段信息和数据结构信息,并与全文信息整合,形成XML格式的标准化数据产品。

3.1.2 标准化图像数据加工

3.1.2.1 获取完整数据源

通过FTP或网络共享或移动硬盘等方式获取中国专利文献单行本原始数据,通过数据包解压、校验及解析等手段验证原始数据包的完整性、可用性和与公报数据的一致性。

3.1.2.2 标准化图像加工

1) 生成PDF文件:对专利全文PDF图像文件进行导航页签分析及规范化处理,为PDF数据设置页标签,标签名称格式为三位大写字母+五位阿拉伯数字页码信息,同时对XML文件的内容缺失和页码标识错误进行纠错。

2) 命名标准化:经标签标准化后的PDF图像文件,按照《专利文献数据规范》(ZC 0014—2012)中的52位标识码要求,利用XML文件中提供的公开号信息进行文件的标准化命名,同时需要与图像数据对应的文本数据进行对比,确保图像与文本采用唯一的标准号(包括申请日、申请号、公开日、公开号)。

3) 存储标准化:经命名标准化的PDF图像文件,按照规范化的存储目录进行转存。

3.1.3 标准化数据导出及分发

1) 对加工完成的标准化文本数据和图像数据按甲方业务要求进行压缩,生成ZIP格式的压缩包,并提供按需定制的标准数据导出和分发服务。

2) 对于中国专利标准化文本数据,可以同时提供符合ZC 0014—2012标准的数据和符合ZC 0012.2—2006标准的数据。

3.2 内部质检

1) 乙方应对加工完成的成品数据进行机器全量质检。

2) 乙方应对每次提交的成品数据进行不低于10%的人工抽检。

3) 对质检不合格的数据批次返回重新加工。

3.3 数据交付

完成加工及质检后,乙方应按照甲方要求的方式交付成品数据。

3.4 数据交付后的数据维护

对于已交付给甲方的数据,如甲方在验收及使用中发现遗漏或错误,无论合同期终止与否,乙方都应免费给予补充或修改。乙方应于0.5个小时之内响应,并于5个工作日内

日内完成补充或修改，并经质检合格后交付。

4、交付物

- 1) 乙方应对加工完成的成品数据进行机器全量质检。
- 2) 乙方应对每次提交的成品数据进行不低于 10% 的人工抽检。
- 3) 对质检不合格的数据批次返回重新加工。

5、技术服务工作报告

乙方应每月向甲方提交上一个月的技术服务工作月报。工作月报应当全面地、具体地体现乙方上一个月的技术服务工作情况，其内容包括但不限于：乙方接收的原始数据的类型和数量、乙方实际完成的数据产品的类型和数量、数据产品质量是否符合合同要求、数据处理和加工的时间周期是否符合合同要求、项目进展等。

6、用户评价

根据甲方的需要，甲方可以要求乙方针对本合同规定的技术服务时间内乙方所提供的的数据产品和服务开展用户评价工作，在甲方规定的时间内向甲方提交用户评价报告。

第三条 技术服务标准

乙方应按照以下标准和规则为甲方提供技术服务：

- 1、《专利数据加工流程及质量管理规范（0002-2012）》；
- 2、《专利数据元素标准第 2 部分：关于用 XML 处理中国发明、实用新型专利文献数据的暂行办法（ZC 0012.2—2006）》；
- 3、《专利文献数据规范（ZC 0014—2012）》；
- 4、《中国专利数据标准化处理及数据集成项目加工及质检规则》。

第四条 技术服务质量

1、准确性

- 1) 乙方保证交付数据的准确性，及时发现数据加工过程中的各种问题并予以纠

正。

2) 专利公报数据(著录项目、名称、摘要和附图等)的标记差错率不超过万分之一(按字符数计算);公开号、公开日、授权公告号、授权公告日、申请号、专利号、申请日、优先权日、优先权号等重要著录项目的差错率不超过万分之一(按项目数计算);图像文件的差错率不超过万分之一(按页数计算);代码化数据文档的标记差错率不超过万分之一(按标记项计算)。

3) 交付数据不得缺项、漏项。

2、及时性

乙方应严格按照如下时限要求完成数据加工工作:

1) 乙方在接收到现档数据后的五个工作日内完成数据的标准化加工工作。

2) 乙方按公报周期(目前每周二和周五)向甲方提交通过内部检测的成品数据。

3) 乙方应在每月10日前向甲方交付上月完成的包括成品数据在内的交付物。

4) 待完成当年的加工数据后,乙方向甲方提交一套当年完整的成品数据。

3、完整性

乙方应保证交付数据的完整性,确保所有数据源均按合同要求进行加工,保证加工后的数据能满足甲方各类业务需要。

第五条 合同价

1、合同总价款

该项目的合同总价款为:人民币捌佰陆拾捌万元整(8,680,000元)。

2、付款安排

1) 合同签订之后,根据甲方时间安排,向乙方支付人民币壹佰肆拾肆万玖仟柒佰陆拾元整(1,449,760元)

2) 根据甲方时间安排,向乙方支付人民币壹佰肆拾肆万玖仟柒佰陆拾元整(1,449,760元)

3) 后续付款按照项目实际执行进度来支付。

3、付款要求

甲方每次付款前,乙方应向甲方开具符合甲方要求的与甲方当次付款金额等额的正

式发票；否则，甲方有权拒绝付款。乙方出示的发票应包含足够详细的内容，使甲方能够确定金额的准确性；如乙方提交的发票不合格，甲方可延迟付款。

4、乙方账户

甲方应将价款支付至下述乙方指定账户内。

账户名称：知识产权出版社有限责任公司

开户银行：工商银行北京北太平庄支行

账号：0200010009004601324

若乙方账户信息发生变更，须提前 20 个工作日书面通知甲方；否则，由此造成延期支付或乙方不能收款的责任由乙方承担。

第六条 考核和验收

1、考核

甲方按照合同要求按月对交付物进行加工时效考核和加工质量考核。

加工时效考核是指乙方应按照时限要求完成当期工作，如逾期视为当月考核不合格。加工质量考核是指交付物应达到质量要求；如果未达到质量要求，乙方应于 5 个工作日内进行整改并提交甲方，如逾期或整改后的交付物仍不能达到质量要求，视为当月考核不合格，乙方仍须继续整改至合格为止。

2、验收

1) 在乙方完成合同约定的全部技术服务工作之后或者在甲方认为必要时，甲方对乙方所交付的数据产品进行验收。乙方向甲方交付的数据产品，必须符合本合同的规定。

2) 乙方应提供经乙方相关负责人签字并加盖公章的验收报告。验收报告应至少包括服务期内乙方详细工作总结，交付的成品数据的数量情况和质量情况，交付

物清单,乙方执行合同义务转让禁止、知识产权、保密等方面要求的情况等内容。

3)如验收未通过,甲方有权暂不签署验收报告并不支付相关款项。乙方应在5个工作日内按照验收意见进行弥补、改进或完善,待甲方重新组织验收并且验收合格,甲方再向乙方支付相关款项。如果乙方在规定的验收日后的一个月,经弥补、改进或完善后仍达不到验收标准的,甲方有权终止合同并拒绝支付相关款项。

第七条 乙方合同义务转让禁止以及乙方保证

1、本合同不接受转包、分包。未经甲方事先书面同意,乙方不得将本合同约定的乙方部分或全部技术服务工作或合同义务,转让给任何第三方承担。

2、乙方保证:乙方根据本合同约定履行义务的行为和过程及乙方交付给甲方的任何工作成果,不存在违反国家法律法规规定、侵犯国家、集体或公众利益或侵犯任何第三人的合法权益的权利瑕疵或情形;甲方根据本合同约定使用乙方交付的任何工作成果,均不受任何第三人的限制,不会被任何第三人提出权利主张或侵权主张,且不侵犯任何第三人的知识产权以及其他合法权益。

第八条 保密约定

1、乙方在为甲方提供服务的过程中获得的有关甲方的任何属于保密协议范围内的信息、数据、资料(不论是甲方提供或披露的还是乙方偶然获得的)以及因本项目履行所取得的任何工作成果,均为保密信息。

2、未经甲方事先书面同意,乙方不得将该等信息自行使用或披露给任何第三方,包括但不限于在著作、会议、教学、媒体、论文等任何公开场合部分或全部引用、发表因履行本项目所获得有关甲方的任何案例、资料、以及因本项目履行所产生的任何工作成果。

3、乙方保证不向本项目下提供服务的人员之外的其他人员披露本项目下的保密信息。乙方应告知并采取必要的有效措施保证参与本项目项下服务项目之乙方人员履行本项目项下的保密义务。若乙方人员违反本项目项下的保密义务,乙方应承担连带保证责任。

4、乙方保证在项目验收后甲方在项目实施过程中提供的相关信息和资料全部返还

1、本合同涉及的各类数据资源及全部工作成果（包括源数据、数据中间产品以及成品数据等）的知识产权属于甲方所有，未经甲方书面授权许可，乙方不得擅自出售或使用，不得许可第三人使用、不得与第三人共同使用。

2、乙方保证其所提供的全部工作成果均没有侵犯任何第三方的知识产权及其它权

益。如甲方因采用乙方的工作成果而引起第三方提起有关侵权之指控、争议或索赔，则乙方应在此等事件发生后的合理时间内以自己的费用解决，保护甲方的利益不受损害，并承担因此给甲方造成的全部损失和费用（包括但不限于法院诉讼的费用、合理的律师酬金和费用、损害赔偿费用、罚款以及其它所有的损失或损害等等）。同时，乙方应自行承担费用采取补救措施，使所提供的工作成果不侵犯他人权利并符合本合同。

在本合同履行过程中，确因外部因素的不可抗力，导致部分或全部数据处理失败，乙方应尽快通知甲方，且须在不可抗力发生后三日内以书面形式向甲方提供详细情况报告及不可抗力对履行本合同的影响程度说明，并提供由公证机关作出的公证证明。在乙方做出上述行为后，甲乙双方共同协商承担风险责任。

在乙方做出上述行为后,甲乙双方共同协商承担风险责任。

1、在任何一批次交付物中，若交付物不符合合同质量要求，甲方可扣除本批次交付物的数据量对应合同金额的 1%作为乙方违约金（根据乙方本款约定违约行为的次数，该违约金可重复计算和扣除）。当违约金发生并达到合同总金额的 10%后，甲方有权终止本合同，并有权要求乙方退回已支付的金额，及赔偿甲方因此遭受的全部经济损失。

2、乙方延误或部分延误交付交付物或服务的，应在延误发生 3 日内以书面形式将延误的事实、可能延误的时间和原因通知甲方，并提供更新的项目计划，阐述由于延

误对整个项目的影响。甲方在收到乙方通知后，将有权选择延长交付时间、和/或收取延误赔偿费、和/或解除合同。乙方每出现一次逾期交付，甲方可延后付款并扣除逾期批次数据量对应合同金额的1%作为乙方违约金（根据乙方本款约定违约行为的次数，该违约金可重复计算和扣除）。当违约金发生并达到合同总金额的10%后，甲方有权终止本合同，并有权要求乙方退回已支付的金额，及赔偿甲方因此遭受的全部经济损失。3、乙方连续2个月考核不合格，甲方有权终止合同。因合同终止而造成的损失由乙方承担全部责任。

4、对于乙方任何违反本合同约定的行为，甲方均有权要求乙方立即停止违约行为

并予以纠正；如乙方拒不纠正、未能在甲方书面要求的纠正期限内纠正或纠正仍不符合本合同约定的，甲方有权终止合同；甲方按照前款前述约定终止合同的，乙方应向甲方支付合同总价款1%的违约金；甲方有权从合同约定的甲方应付款中直接扣除；乙方前述违约行为，如给甲方造成损失，乙方还应赔偿由此给甲方造成的全部损失。

5、甲方终止合同的，甲方立即对乙方截至合同终止之日实际完成的数据产品进行验收，以甲方实际验收合格的数据产品数量为依据计算和支付费用；如果甲方已付给乙方的费用多于乙方根据本款约定最终实际应得的费用，则乙方应自前述甲方多支付部分的费用金额确定之日起2日内向甲方足额退还。

6、甲方终止全部或部分合同的，甲方有权按照其认为适当的条件和方法向任何第三方购买与乙方未提交的可交付物类似的产品或服务，乙方应承担甲方为此所支付的额外费用。就合同中未终止部分，乙方应继续执行。

7、甲方按本条规定要求乙方承担责任外，并不影响甲方继续以其它任何形式要求乙方承担违约责任和赔偿。

8、本合同所称甲方损失，包括但不限于：甲方直接损失、间接损失、第三人（包括但不限于：合法权益被侵害人）向甲方索赔款项以及甲方因调查乙方违约行为及向乙方主张权利而支出的调查费、差旅费、误工费、律师费、诉讼费及其它全部合理费用。

第十二条 合同变更或解除

1、变更

除本合同另有约定外，本合同的任何变更，均须经甲乙双方协商一致，并以书面形式

式确定。

但有下列情形之一的，一方可以向另一方提出变更合同的请求，另一方应当在 30 个工作日内予以答复；逾期未予答复的，视为同意：

1) 甲乙双方协商一致的。

2) 订立本合同所依据的客观情况发生重大变化，致使合同无法履行的。

2、解除

1) 在本合同执行过程中，如果因国家财政政策变化或财政预算调整导致本合同无

法继续执行，则本合同无条件自动终止。

2) 除本合同另有约定外，本合同的解除须经甲乙双方协商一致并书面形式确定。

第十三条 争议解决

本合同履行过程中所发生的一切争议，甲乙双方应通过友好协商解决；协商不成，应提交国家知识产权局进行调解；调解不成，任何一方均应向甲方住所地人民法院提起诉讼解决。

第十四条 合同附件

本合同所附下列文件是构成本合同不可分割的部分：

附件：《中国专利数据标准化处理及数据集成项目加工及质检规则》。

第十五条 合同生效

本合同经甲乙双方法定代表人或授权代表签字并加盖双方公章或合同专用章之日

起正式生效。在甲乙双方全部执行完本合同规定的内容后，本合同自然终止。

第十六条 联系方式

甲 方	名称（或姓名）	国家知识产权局专利局		
	授权代表	钱红缨		
	联系人	李耐		
	通讯地址	北京市海淀区蓟门桥西土城路6号		
	联系人电话	010-62083734	传真	
乙 方	名称（或姓名）	知识产权出版社有限责任公司		
	法定代表	诸敏刚		
	联系人	路志明		
	通讯地址	北京市海淀区气象路50号院		
	联系人电话	010-82000860-8218	传真	

第十七条 其他约定

甲乙双方约定本合同其他相关事项为：

- 1、乙方应当保证其交付给甲方的技术服务和合同产品不侵犯任何第三方的合法权益。如发生第三方指控甲方实施的技术侵权的，乙方应当承担全部责任。
- 2、未在本合同中约定但与本合同履行有关的技术标准和规则，由甲乙双方或包括

甲方(盖章): 国家知识产权局
法定/授权代表人(签字): 徐政刚
日期: 2017年12月1日

乙方(盖章): 中国出版集团有限责任公司
法定/授权代表人(签字): 徐政刚
日期:

(以下无正文)

4、本合同一式捌份, 甲方执肆份, 乙方执肆份, 每份均具有同等法律效力。

3、本合同未尽事宜, 双方另行协商并签订书面补充协议明确。

甲方指定的第三方共同商定。

中国专利数据标准化加工服务项目加工规则

1 根据专利文献数据规范制定文本数据加工规则

1.1 通用属性及元素规则

通用数据类型可能的取值列表如下：

表3 数据类型可能的取值列表

数据类型	说明
字符串 (string)	通过字符串形式表达值的类型。
数字型 (number)	通过从“0”到“9”数字形式的值的类型。
日期型 (date)	通过YYYYMMDD形式表达的值类型，符合GB/T 7408。
日期时间型 (datetime)	通过YYYYMMDDThhmmss形式表达的值类型，符合GB/T 7408。
布尔型 (boolean)	两个且只有两个表明条件的值，如on/off、True/False。
二进制 (binary)	上述无法表示的其他数据类型，比如图像、音频等。

1、基础类主要元素

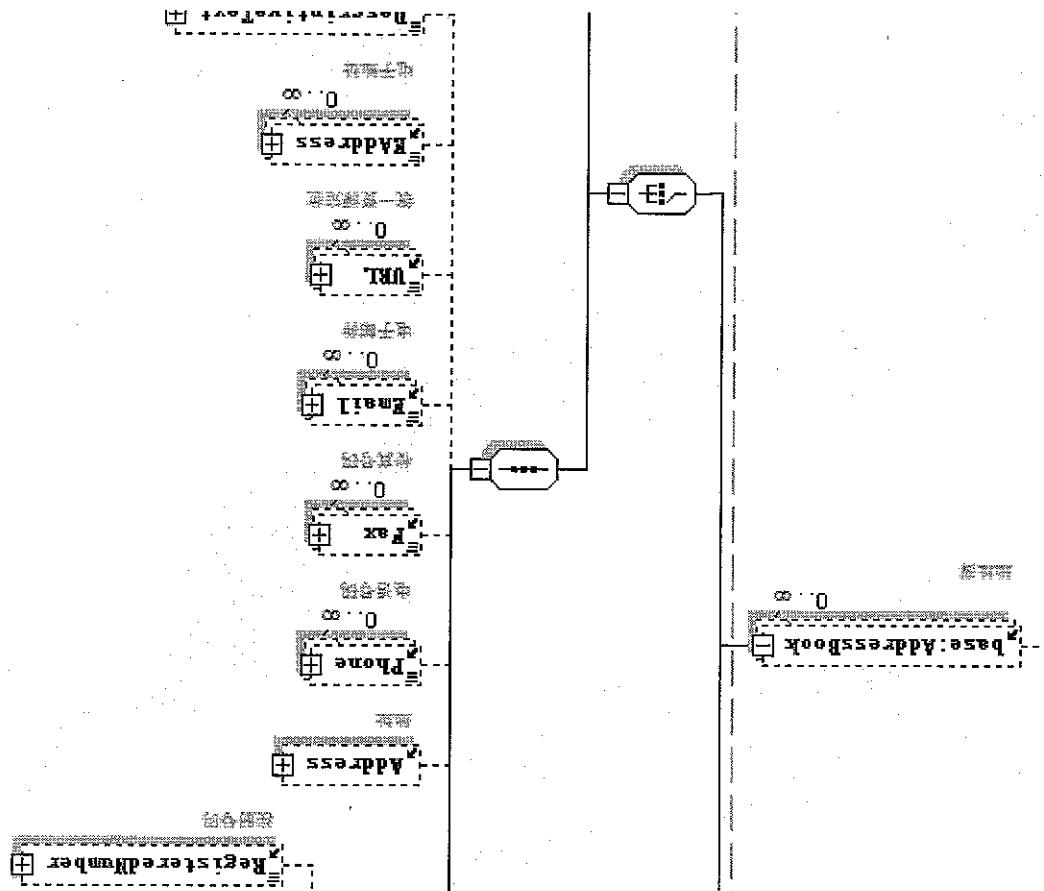
A.1 元素类别的名称和编号列表

元素类别	中文名称	英文名称	元素类别编号
基础类 BA	时间	Date/Time	0100
	地址簿	AddressBook	0200
	地址	Address	0201
	名称	Name/NameGroup	0202
	表格	Tables	0300
	图像	Images	0400
	文档页	DocPage	0500
	化学	Chemistry	0600
	数学	Maths	0700
	列表	Lists	0800
	文档结构	DocumentStructure	0900
	字体的格式	Format	1000
	引用文献的描述信息	Citation	1100
	文献标识	DocumentID	1200
	签名	Signature	1300
	其它	Other	1400
	数值范围	NumericalRange	1500

基础元素 Date:

标识符		BA0100001
名称	中文名称	日期
	英文名称	Date
定义		某个特定日期，日期的具体含义要应用上下文而定
子元素		无
属性		无
值域		日期格式的字符串
数据类型		string
Schema 代码		<code><xs:element name="Date" type="xs:string"></code> <code></xs:element></code> <code><xs:annotation></code> <code></xs:annotation></code> <code><xs:documentation>日期</xs:documentation></code> <code></xs:element></code>
结构图		
备注		

基础标签 Addressbook:





1. $\frac{1}{2}$

标识符		BA0200001	
名称	中文名称	地址簿	
	英文名称	AddressBook	
定义		用于联系个人或组织的地址、名称等信息	
子元素		AddressBuilding, AddressCity, AddressFloor, AddressLine, AddressMailCode, AddressRoom, County, PostBox, PostCode, State, Street, Text, ZIP0513Code	
属性		id, lang	
值域		属性 id 表示标识号, 可选。 属性 lang 表示语言代码, 可选。	
数据类型			
Schema 代码		<pre><xs:complexType name="AddressBookType"> <xs:choice> <xs:sequence> <xs:sequence> <xs:choice></pre>	

B.1.12 BA0201001/Address

标识符		BA0201001
名称	中文名称	地址
	英文名称	Address
定义		地址的相关信息
子元素		AddressBuilding, AddressCity, AddressFloor, AddressLine, AddressMailCode, AddressRoom, County, PostBox, PostCode, State, Province, Street, Text, ZIPCode
属性		无
值域		
数据类型		
Schema 代码		<xs:complexType name="AddressType"> <xs:sequence>

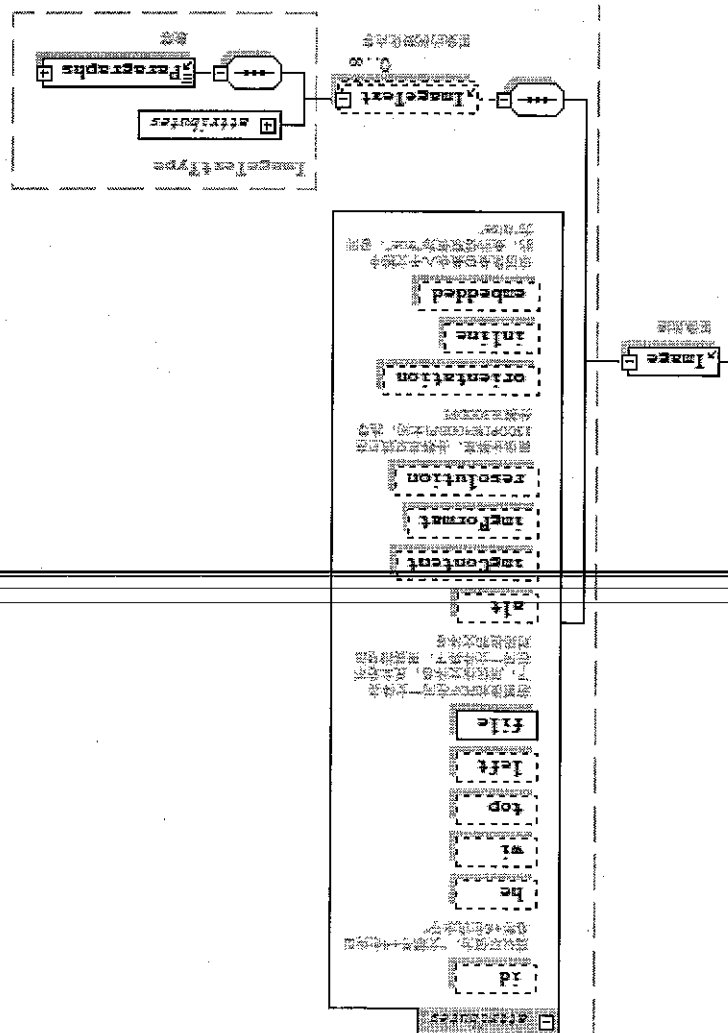
基础标签 Table:

B.1.46 BA0300002/Table

标识符		BA0300002
名称	中文名称	表格
	英文名称	Table
定义		表格, 遵循 TableML 加工标准
子元素		无
属性		id

值域	文本字符串描述的表格内容
数据类型	string
Schema 代码	<xs:complexType name="TableType"> <xs:sequence>

基础标签 Image:



标识符	名称		定义	子元素	属性	值域
	中文名称	英文名称				
BA0100003	图像数据	Image	图像数据可以是绘画、图形或者复来的元素（数学式、表格、化学式）等	ImageText	id, he, wi, top, left, file, alt, imgContent, imgFormat, orientation, inline, embedded, resolution	若图像和 PDF 在同一文件夹下，则属性 file 仅指文件名；反之，若不在同一文件夹下，属性 file>需要指明绝对路径和文件名，必填。 属性 id 为标号，建议取值为：“文献号+4 位类型名称+6 位流水号”，可选。

		属性 he 表示图像高度，建议以毫米为单位，可选。 属性 wi 表示图像宽度，建议以毫米为单位，可选。 属性 top 表示图像上角点坐标，建议以毫米为单位，可选。 属性 left 表示图像左角点坐标，建议以毫米为单位，可选。 属性 alt 表示在图像无法显示时的替代文本，可选。 属性 imgContent 表示图像内容，取值包括 "photograph", "dne", "drawing", "character" 和 "undefined"，可选。 属性 imgFormat 表示图像格式，取值包括 "pdf", "tif", "t33", "t35", "可选。 属性 orientation 表示图像的方向，取值包括 "landscape" 和 "portrait"，默认值为 "portrait"，可选。

		属性 embedded, 图像是否嵌入文档，“是”取值 "true", “否”取值 "false", 可选。 属性 resolution, 图像分辨率，分辨率取值应在 150DPI 和 400DPI 之间，推荐分辨率 300DPI。

基础元素 lang:

lang, 两位小写字母，符合 ISO639-1,

aa	Afar	阿尔法语
ab	Abkhazian	阿布哈西亚语
af	Afrikaans	南非荷兰语
am	Amharic	阿姆哈拉语
ar	Arabic	阿拉伯语
as	Assamese	阿萨姆语
ay	Aymara	艾玛拉语
az	Azerbaijani	阿塞拜疆语
be	Byelorussian	白俄罗斯语
bg	Bulgarian	保加利亚语
bh	Bihari	比哈里语
bi	Bislama	比斯拉马语
bn	Bengali; Bangla	孟加拉语
bo	Tibetan	藏语
br	Breton	凯尔特语
ca	Catalan	加泰罗尼亚语
co	Corsican	柯西嘉语
cs	Czech	捷克语
cy	Welsh	威尔士语
da	Danish	丹麦语
de	German	德语
dz	Bhutan	不丹语
el	Greek	希腊语
en	English	英语
eo	Esperanto	世界语
es	Spanish	西班牙语
et	Estonian	爱沙尼亚语
eu	Basque	巴斯克语
fa	Persian	波斯语
fi	Finnish	芬兰语
fj	Fiji	斐济语
fo	Faeroese	法罗语
fr	French	法语
fy	Frisian	弗里西亚语 (或弗里斯兰语)
ga	Irish	爱尔兰语
gd	Scotts Gaelic	苏格兰盖尔语
gl	Galician	保加西亚语
gn	Guarani	瓜拉尼语
ha	Hausa	豪萨语
aa	Afar	阿尔法语
ab	Abkhazian	阿布哈西亚语
af	Afrikaans	南非荷兰语
am	Amharic	阿姆哈拉语
ar	Arabic	阿拉伯语
as	Assamese	阿萨姆语
ay	Aymara	艾玛拉语
az	Azerbaijani	阿塞拜疆语
be	Byelorussian	白俄罗斯语
bg	Bulgarian	保加利亚语
bh	Bihari	比哈里语
bi	Bislama	比斯拉马语
bn	Bengali; Bangla	孟加拉语
bo	Tibetan	藏语
br	Breton	凯尔特语
ca	Catalan	加泰罗尼亚语
co	Corsican	柯西嘉语
cs	Czech	捷克语
cy	Welsh	威尔士语
da	Danish	丹麦语
de	German	德语
dz	Bhutan	不丹语
el	Greek	希腊语
en	English	英语
eo	Esperanto	世界语
es	Spanish	西班牙语
et	Estonian	爱沙尼亚语
eu	Basque	巴斯克语
fa	Persian	波斯语
fi	Finnish	芬兰语
fj	Fiji	斐济语
fo	Faeroese	法罗语
fr	French	法语
fy	Frisian	弗里西亚语 (或弗里斯兰语)
ga	Irish	爱尔兰语
gd	Scotts Gaelic	苏格兰盖尔语
gl	Galician	保加西亚语
gn	Guarani	瓜拉尼语
ha	Hausa	豪萨语

hi	Hindi	印地语
hr	Croatian	克罗地亚语
hu	Hungarian	匈牙利语
hy	Armenian	亚美尼亚语
ia	Interlingua	拉丁国际语
is	Icelandic	印度尼西亚语
it	Italian	冰岛语
iw	Hebrew	意大利语
ja	Japanese	现代以色列语 (希伯来语)
ji	Yiddish	日语
jw	Javanese	意第绪语 (依地语)
ka	Georgian	爪哇语
kk	Kazakh	格鲁吉亚语
kl	Greenlandic	哈萨克语
km	Cambodian	格陵兰语
kn	Kannada	柬埔寨语
ko	Korean	卡纳达语
ks	Kashmiri	朝鲜语
ku	Kurdish	克什米尔语
ky	Kirghiz	库尔德语
la	Latin	吉尔吉斯语
ln	Lingala	拉丁语
lo	Laotian	林加拉语
lv	Latvian, Lettish	立陶宛语
mg	Malagasy	拉脱维亚语, 列托语
mi	Maori	马尔加什语
mk	Macedonian	毛利语
ml	Malayalam	马拉雅拉姆语
mn	Mongolian	蒙古语
mo	Moldavian	摩尔达维亚语
mr	Marathi	马拉地语
ms	Malay	马来语
mt	Maltese	马耳他语
my	Burmese	缅甸语
na	Nauru	瑙鲁语
ne	Nepali	尼泊尔语
nl	Dutch	荷兰语
no	Norwegian	挪威语
oc	Occitan	奥坎语
pl	Polish	旁遮普语

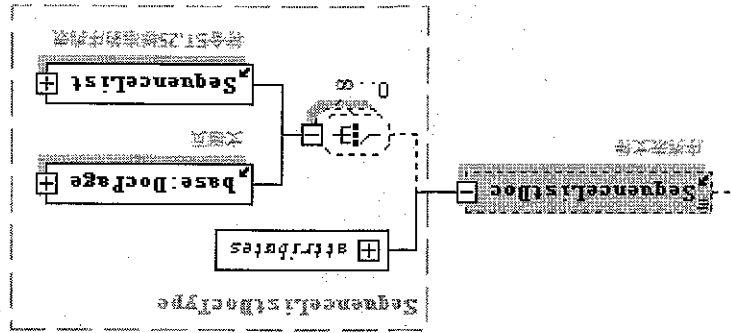
复杂标签 WOAendedClaims:

2、复杂类主要元素

属性 status, 状态, 取值如 C (新建)、R (替代)、D (删除)、A (修改) B (过档)

基础元素 status:

zu	Zulu	祖鲁语
zh	Chinese	汉语
yo	Yoruba	约鲁巴语
xh	Xhosa	科萨语
wo	Wolof	沃洛夫语
vo	Volapuk	沃拉普克语
vi	Vietnamese	越南语
ur	Urdu	乌尔都语
uk	Ukrainian	乌克兰语
tw	Twi	阿维尼语
tt	Tatar	鞑靼语
tr	Turkish	土耳其语
to	Tonga	汤加语
tn	Setswana	茨瓦纳语
tl	Tagalog	他加禄语
tk	Turkmen	土库曼语
ti	Tigrinya	提格里尼亚语
th	Thai	泰语
tg	Tajik	塔吉克语
ta	Tamil	泰米尔语
sw	Swahili	斯瓦希里语
sv	Swedish	瑞典语
su	Sundanese	巽他语
st	Sesotho	塞索托语
ss	SiSwati	斯瓦特
sr	Serbian	塞尔维亚语
sq	Albanian	阿尔巴尼亚语
so	Somali	索马里语
sn	Shona	绍纳语
sm	Samoan	萨摩亚语
sl	Slovenian	斯洛文尼亚语
sk	Slovak	斯洛伐克语
si	Singhalese	僧伽罗语
sh	Serbo-Croatian	克罗地亚语
sd	Sindhi	信德语
sa	Sanskrit	梵语
ru	Russian	俄语
ro	Romanian	罗马尼亚语
rn	Kirundi	基隆迪语
rm	Rhaeto-Romance	罗曼语
qu	Quechua	盖丘亚语
pt	Portuguese	葡萄牙语
ps	Pashto, Pushto	波兰语



复杂标签 SequenceListDoc:

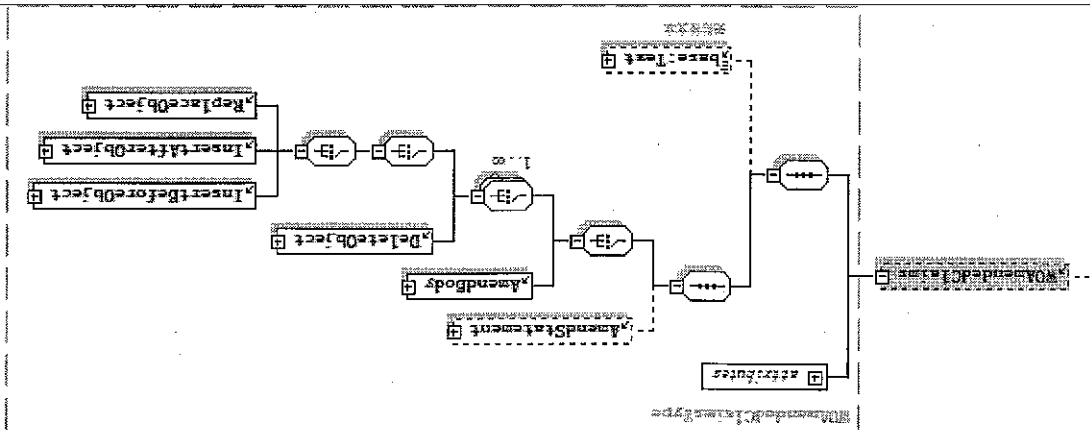
该元素是原始申请的补正信息，内容包括补正声明（AmendStatement）、补正体（AmendBody）或者对对象的操作，操作包括：删除对象（DeleteObject）、对象前插入新内容（InsertBeforeObject）、对象后插入新内容（InsertAfterObject）、替换对象（ReplaceObject）。*

元素 AmendStatement，依据 PCT 19.1 PCT Rule46.4 修改的声明，可以是文档格式式的补正声明，也可以是文本格式的补正声明，分别对应于 AmendStatement/DocPage 和 AmendStatement/Statement。*

元素 AmendBody，补正体可以是文档页格式，存储于 AmendBody/DocPage，也可以是严格规范化的信息，包含说明书 Description，权利要求 Claims，摘要 Abstract，说明书附图 Drawings。这些元素的具体使用和说明见其他部分。*

对于补正的信息还可以通过操作内容体现，例如删除、插入、替换某内容。除删除操作，其他操作要包含操作对象、操作位置和新内容。删除操作 DeleteObject 只需要指出删除位置 and 删除对象，存储于 DeleteObject/@Location，DeleteObject/@Id。而 InsertBeforeObject、InsertAfterObject、ReplaceObject 还需要指出新内容，新内容可以是发明名称、标题、段落、权利要求、图例的一种。以 InsertBeforeObject 为例，新内容存储于 InsertBeforeObject/InventionTitle，InsertBeforeObject/Heading，InsertBeforeObject/Paragraphs，InsertBeforeObject/Claim，InsertBeforeObject/Figure。*

~~B-3-22 NO4-memo/dc|a|msv~~



B.3.27 SequenceListDoc

该元素是序列表文件，包含文档页格式 (SequenceListDoc/DocPage) 和符合 ST.25 标准的序列表 (SequenceListDoc/SequenceList)。前者用于图像格式的序列表，后者用于存储规范化、标准化的序列表。

元素 SequenceList 是符合 ST.25 标准的序列表，内容包含序列表和序列表电子文件中的著录项目 SequenceListBibliographicData，序列个数 SequenceCount，软件版本

SoftwareVersion，序列内容 SequenceContent。

属性 id，标识；

属性 file，序列文件存放的相对路径；

属性 lang，语言代码，遵循 ISO639-1，两位小写字母；

属性 status，状态，取值为过档/新增/更正/删除；

属性 carrier，序列表载体，取值为内部电子件/外部电子件/纸件；

属性 sequenceFileType，序列表文件类型；

属性 electronicFormatType，序列表的格式，取值为 text, pri, other 等；

属性 partOfApplication，表明该序列表是否是申请文件的一部分，取值为 yes, no；

元素 SequenceListBibliographicData，包含申请人信息，发明名称，案卷参考号，

申请信息和优先权信息，分别存储于 SequenceListBibliographicData/Applicant，

SequenceListBibliographicData/InventionTitle

SequenceListBibliographicData/FileReference，SequenceListBibliographicData/

ApplicationReference，SequenceListBibliographicData/PriorityDetails。

元素 SequenceContent 涉及序列内容，包括：序列标识号

SequenceContent/SequenceID，序列长度 SequenceContent/SequenceLength，序列类型

SequenceContent/SequenceType，生物体名称 SequenceContent/Organism，序列特征

SequenceContent/SequenceFeatureDetails，序列公开信息

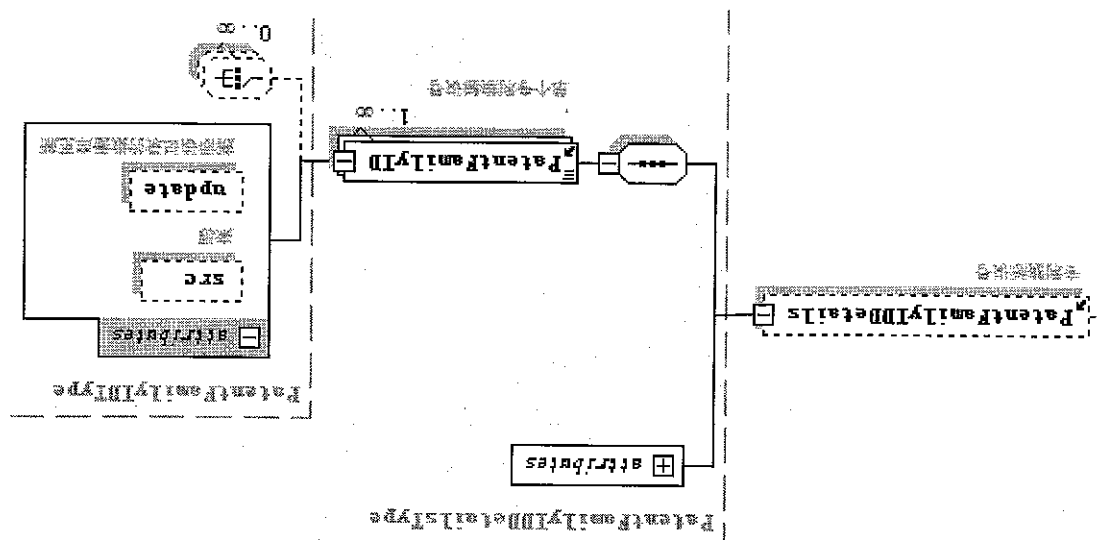
SequenceContent/SequencePublicationInformation，序列描述

SequenceContent/SequenceDescription。除 SequenceContent/SequenceFeatureDetails

和 SequenceContent/SequencePublicationInformation 外，其他元素结构一致，都为字符

串格式，包含属性 id：标识。

复杂标签 PatentFamilyDetails



元素 PatentFamily 用来描述专利族及其成员的相关信息，至少包含专利族标识号 PatentFamilyID 和专利族成员信息 PatentFamily/MemberPatentDetails。在使用 PatentFamily 时，配合其不同属性值，表达专利族各类信息，具体说明如下：⁴

属性 src 表示数据来源，可选；⁴

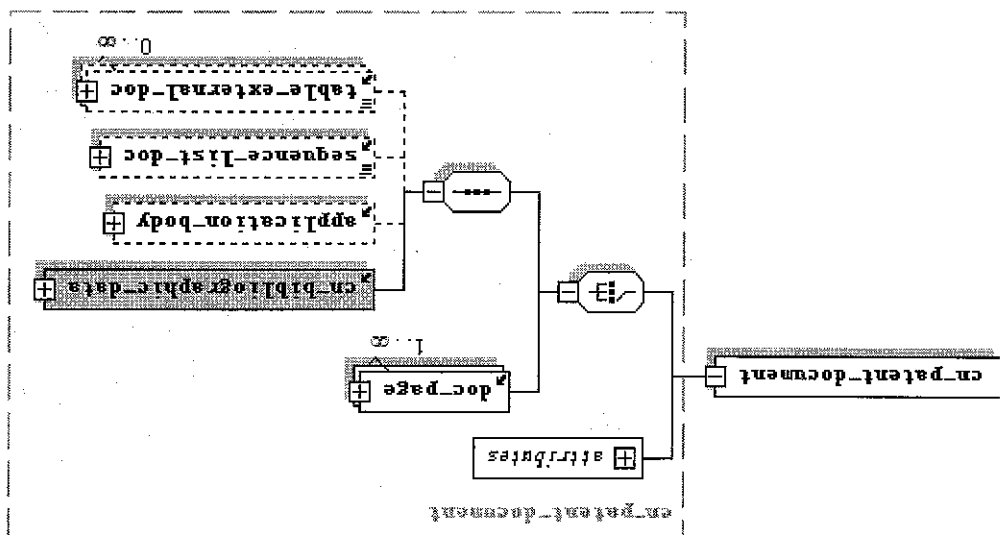
属性 patentDataResourceID 表示专利数据资源标识码，可选；⁴

属性 familyMemberCount 表示。⁴

1.2 著录项目加工规则

1、中国专利基础数据著录项目结构

• DTD-Schema 结构如下：



展开后如下：

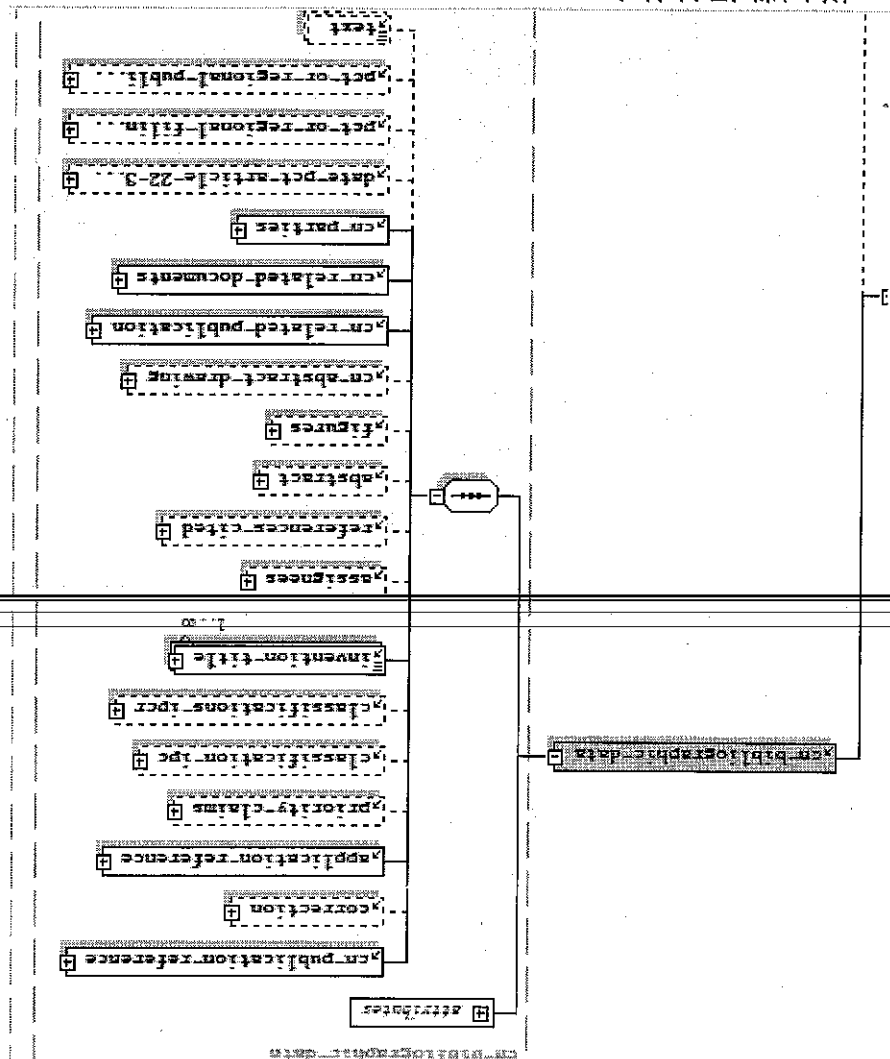
展开后如下:

```

<cn-patent-document lang="zh" country="CN">
  <cn-bibliographic-data>
    <application-body lang="zh" country="CN">
      </cn-patent-document>

```

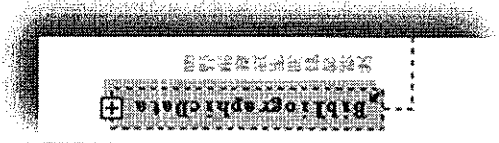
• 样例数据结构如下:



展开后如下:

<business:PatentDocumentAndRelated xmlns:base="http://www.sipo.gov.cn/XMLSchema/business" xmlns="http://oasis-open.org/specs/soextblx" xmlns:xsi="http://www.sipo.gov.cn/XMLSchema/business/DTD/ http://www.sipo.gov.cn/XMLSchema/business/DTDS/CN102015000249243CN00001048427200BFUZH20160106C documentNumber="104842720" kind="B" datePublication="business:PublicationDate">

● 样例数据结构如下:



- DTD-Schema 结构如下:

2、中国标准化数据著录项目结构

```
<cn-parent-document lang="zh" country="CN">
  <cn-bibliographic-data>
    <publication-reference>
      <application-reference>
        <priority-claims>
          <classifications-ipc>
            <invention-title>波导管缝隙天线以及使用该波导管缝隙天线的警报系统</invention-title>
            <abstract>
              <cn-abstract-drawing>
              <cn-related-publication />
              <cn-related-documents />
              <cn-parties>
                <date-pct-article-22-39-fulfilled>
                <date>20151123</date>
              </date-pct-article-22-39-fulfilled>
            </date-pct-article-22-39-fulfilled>
          </date-pct-article-22-39-fulfilled>
        </date-pct-article-22-39-fulfilled>
      </publication-reference>
    </cn-bibliographic-data>
  </cn-parent-document>
</cn-parent-document>
```


中国基础数据 XML 样例如下:

on-publication-ref: 公开信息		PublicationReference		加工规则 (无需转换则取原值)	
id		id		无值时不出现	2位小写字母
lang		lang		转为小写字母	2位小写字母
status		status		不取原值	2位大写字母
country		country			2位大写字母
dataFormat		dataFormat		dataFormat='standard', 必须出现;	2位大写字母
sequence		sequence		取值1	必须
base: DocumentID		base: DocumentID			
ATT: lang		ATT: lang		转为小写字母	2位小写字母
ATT: lang		ATT: lang		转为小写字母	2位小写字母
WIP0513Code		WIP0513Code		需要标准化转换	2位大写字母
DocNumber		DocNumber			
Kind		Kind			
Name		Name			
ATT: nameType		ATT: nameType			
Date		Date			8位数字
PublicationReference		PublicationReference			
id		id			
lang		lang		转为小写字母	2位小写字母
status		status			
country		country			2位大写字母
dataFormat		dataFormat		dataFormat='original', 必须出现;	2位大写字母
sourceDB		sourceDB		national office	
sequence		sequence		取值1	必须
base: DocumentID		base: DocumentID			
document-id		document-id			
ATT: lang		ATT: lang		转为小写字母	2位小写字母
country		country			2位大写字母
doc-number		doc-number		取中国库中的 原始公开号	只含数字
kind		kind			
name		name			
ATT: name-type		ATT: name-type			
date		date			8位数字
gazette-reference		gazette-reference			
id		id			
country		country			2位大写字母
lang		lang		转为小写字母	2位小写字母
gazette-num		gazette-num			
日期		日期			8位数字
文本		文本			

标准化成品文本 XML 数据中, 标签 'PriorityDetails' 将出现多次, 每一条优先权信息都包含原始格式和标准化格式, 由属性 'dataFormat' 的取值控制, 如 dataFormat="standard" 表示标准格式, dataFormat="original" 表示原始格式; 优先权信息包含的主要元素有优先权国家、优先权号码和优先权日期; 优先权信息整体结构和层级对应关系及加工规则如下:

3、优先权信息规则如下:

```

<business:ApplicationReference appl-type="发明专利">
  <document-id>
    <country>CN</country>
    <doc-number>201510767458.4</doc-number>
    <date>20151111</date>
  </document-id>
  <application-reference>
    <business:DocumentID>
      <base:DocumentID>
        <base:WIPOST3Code>CN</base:WIPOST3Code>
        <base:DocNumber>201510767458.4</base:DocNumber>
        <base:Date>20151111</base:Date>
      </base:DocumentID>
    </business:DocumentID>
    <business:ApplicationReference appl-type="10" dataFormat="original" sequence="1" sourceDB="national office">
      <business:DocumentID>
        <base:DocumentID>
          <base:WIPOST3Code>CN</base:WIPOST3Code>
          <base:DocNumber>102015000761458</base:DocNumber>
          <base:Date>20151111</base:Date>
        </base:DocumentID>
      </business:DocumentID>
    </business:ApplicationReference>
  </business:ApplicationReference>

```

对应的中国标准化成品 XML 数据样例如下:

```

<application-reference appl-type="发明专利">
  <document-id>
    <country>CN</country>
    <doc-number>201510767458.4</doc-number>
    <date>20151111</date>
  </document-id>
  <application-reference>

```

priority claims		所有优先权数据		PriorityData	加工规则（无需转换则取原值）	
		优先权数据	Priority			
		序号	优先权数据的顺序	sequence	必选：有原值取原值，无原值按多值出现顺序排列 sequence=1； sequence=2； sequence=3...	“优先权国家与公开国家相同：优先权国家与公开国家不同：international
		文献种类	kind	ATI		
					dataFormat	dataFormat='standard'
		优先权国家	WIPOST3Code			
		优先权号	DocNumber			需要标准化转换
		优先权日期	Date			
		在先申请的提交局	FilingOffice			
		地区	Region			
		国别代码	WIPOST3Code			
		国别代码	WIPOST3Code			
		国别代码	WIPOST3Code			
		要求受理局向国际递交在先申请优先权证明和文件，仅对PCT申请和申请一起提交的优先权文件	PriorityDocRequested			应值 取原始格式的对
			PriorityDocAttached			

对应的中国标准化成品XML数据样例如下:

```
<cn-patent-document lang="zh" country="CN">
  <cn-bibliographic-data>
    <cn-publication-reference>
      <application-reference appl-type="发明专利">
        <priority-claims>
          <priority-claim sequence="1" kind="international">
            <country>JP</country>
            <doc-number>2013-127613</doc-number>
            <date>20130618</date>
          </priority-claim>
          <priority-claim sequence="2" kind="international">
            <country>JP</country>
            <doc-number>2014-040325</doc-number>
            <date>20140303</date>
          </priority-claim>
        </priority-claims>
      </application-reference>
    </cn-publication-reference>
  </cn-bibliographic-data>
</cn-patent-document>
```

中国基础数据XML样例如下:

[illegible]

标准化成品文本XML数据中，IPCR分类信息标签‘ClassificationPCRDetais’可以是多值，原始数据分类号数量与标准化后的分类号数量一致，此标签主要的加工及拆分对象为分类号信息，依据分类号信息的总字符长度、位置和内容拆分成标准化的分类信息格式，并保留原始格式存放于‘ClassificationPCR/Text’标签内，具体对应规则如下：

4、IPCR分类信息规则如下：

```
<business:PriorityDetails>
<business:Priority kind="international" dataFormat="standard" sequence="1">
<base:WIPOST3Code>JP</base:WIPOST3Code>
<base:DocNumber>002013000127613</base:DocNumber>
<base:Date>20130618</base:Date>
</business:Priority>
<business:Priority kind="international" dataFormat="standard" sequence="2">
<base:WIPOST3Code>JP</base:WIPOST3Code>
<base:DocNumber>102014000040325</base:DocNumber>
<base:Date>20140303</base:Date>
</business:Priority>
<business:Priority kind="international" dataFormat="original" sourceDB="national office" sequence="1">
<base:WIPOST3Code>JP</base:WIPOST3Code>
<base:DocNumber>2013-127613</base:DocNumber>
<base:Date>20130618</base:Date>
</business:Priority>
<business:Priority kind="international" dataFormat="original" sourceDB="national office" sequence="2">
<base:WIPOST3Code>JP</base:WIPOST3Code>
<base:DocNumber>2014-040325</base:DocNumber>
<base:Date>20140303</base:Date>
</business:Priority>
```

classifications-ipc	第8版IPC 的版分类 号信息	ClassificationIPCDetail	凡IPC标准化规则
ATTid	id	id	
id	id	id	
classification-ipc	第8版IPC 的版分类	ClassificationIPC	初加工取 值original (表示原始)
ipc	id	id	
sequence	sequence	sequence	
ipc-version- indicator	第8版IPC 的版本标 记, 用日 期表示		
date	日期	IPCVersionDate	
classification-level	分类级别	ClassificationLevel	
section	部	Section	
class	大类	MainClass	
subclass	小类	Subclass	
main-group	大组	MainGroup	
subgroup	小组	Subgroup	
symbol-position	分类号位	SymbolPosition	
classification-value	分类的属 性	ClassificationValue	
action-date	生效日期	base: ActionDate	
date	日期	date	
generating-office	生成局	GeneratingOffice	
country	国别	base: WIPO30	无值时填写CN
classification-status	分类状态	ClassificationStatus	
classification-data-source	国际分类 数据来源	ClassificationDataSource	无值时填写H
text	不能细分 的信息	base: text	

IPCR规则分析					
位长	数量	拆分规则	加工规则	样例	备注
18	3710707	1位字母 (A~H) +2位数字 +1位字母 +1~3位数字 +1位 '/' +2~n位数字 +1位 '(' +1位年代	部<Section>=1位字母 (A~H) 大类<MainClass>=2位数字 小类<Subclass>=1位字母 大组 <MainGroup>=1~3位数字(第5位至 '/' 前的数字) 小组<Subgroup>=2~n位数字 ('/' 后至 ' ' 前的数字) IPC版本号 <IPCVersionDate>= ' ' 和 ' ' 中间的年月,需转换为8位日期,一般转为 "20060101" 分类的属性 <ClassificationValue>=最后1位字母 (N/I)	H01H50/02 (2006.01) G02F1/133 (2006.01) A61K39/395 (2006.01) G02F1/1335 (2006.01) G11B21/02 (2006.01) I	
19	3262072			A61K39/395 (2006.01) G02F1/1335 (2006.01)	
17	1236509			C12N9/96 (2006.01)	
20	512176			A61K31/7048 (2006.01)	
21	58873			A61K31/4164 (2006.01) I	
22	24	+1位字母 (N/I) +1位 '/' 信息 (01~12) +2位月份 +1位 ',' +1位 '.' 后) (2006之 信息 (2006之 IPC版本号 <IPCVersionDate>= ' ' 和 ' ' 中间的年月,需转换为8位日期,一般转为 "20060101" 分类的属性 <ClassificationValue>=最后1位字母 (N/I)	F21W1 (E04B1/98 (2006.01) I) 1111 扉页中无错,可将后面 '1111' 35(20) 直接删除; 2. CN101697872A (A61B5/02057 (2006.01) I) 扉页中即为错误数据。		

中国基础数据 XML 样例如下:

```
<classifications-ipcr>
<classification-ipcr>
<text>H01Q 13/22 (2006.01)</text>
</classification-ipcr>
<classification-ipcr>
<text>H01Q 21/06 (2006.01)</text>
</classification-ipcr>
</classifications-ipcr>
```

对应的中国标准化成品 XML 数据样例如下:

中国基础数据 XML 样例如下:

Invention-title		发明名称		InventionTitle		加工规则（无需转换则取原值）	
ATT	id	id号		ATT	id		
	lang	语言代码			lang	转为小写字母，取zh	
				creator	必须出现；		
				dataFormat	必须出现；取值original等		
				sourceDB	必须出现；数据来源取值如national office等		
				processingType	因多合一，初加工取原值original（表示原始）		
b		粗体		Bold			
i		斜体		Italic			
u		在下部的修饰		Underline			
sup		上标		Sup			
sub		下标		Sub			

5、发明名称信息的整体结构和层级对应关系及加工规则如下:

```
<business:ClassificationPCRDetais creator="03" processingType="original">
<business:ClassificationPCR sequence="1">
<business:IPCVersionDate>20060101</business:IPCVersionDate>
<business:Section>H</business:Section>
<business:MainClass>01</business:MainClass>
<business:Subclass>Q</business:Subclass>
<business:MainGroup>13</business:MainGroup>
<business:Subgroup>22</business:Subgroup>
<business:GeneratingOffice>
<base:WIPOST3Code>CN</base:WIPOST3Code>
<business:GeneratingOffice>
<business:Subgroup>06</business:Subgroup>
<business:MainGroup>21</business:MainGroup>
<business:Subclass>Q</business:Subclass>
<business:MainClass>01</business:MainClass>
<business:Section>H</business:Section>
<business:IPCVersionDate>20060101</business:IPCVersionDate>
<business:ClassificationPCR sequence="2">
<business:IPCVersionDate>20060101</business:IPCVersionDate>
<business:Section>H</business:Section>
<business:MainClass>01</business:MainClass>
<business:Subclass>Q</business:Subclass>
<business:MainGroup>21</business:MainGroup>
<business:Subgroup>06</business:Subgroup>
<business:GeneratingOffice>
<base:WIPOST3Code>CN</base:WIPOST3Code>
<business:GeneratingOffice>
<business:Subgroup>06</business:Subgroup>
<business:MainGroup>21</business:MainGroup>
<business:Subclass>Q</business:Subclass>
<business:MainClass>01</business:MainClass>
<business:Section>H</business:Section>
<business:IPCVersionDate>20060101</business:IPCVersionDate>
<business:ClassificationPCR>
<business:ClassificationPCR>
<business:ClassificationPCRDetais>
```

<Invention-title>一种镁合金表面处理办法</invention-title>

对应的中国标准化成品XML数据样例如下:

```
<business:InventionTitle lang="en" dataFormat="original" sourceDB="CPFA" processingType="translation" creator="05">Magnesium alloy surface treatment method</business:InventionTitle>
```

6、专利权人信息的整体结构和层级对应关系及加工规则如下:

assignees		所有专利权人信息		AssigneeDetails		加工规则(无 需转换则取原 值)	
assignee	专利权人	Assignee		Assignee			
		AT	sequence	sequence=1; 顺序排列 值按多值出现 取原值,没原 值选:有原值 必须出现	sequence=2...	必须出现;	
			creator	必须出现;		必须出现;	
			lang	必须出现;		必须出现;	
			dataFormat		必须出现;取 值original等	必须出现;数 据来源取值 national office等	
			sourceDB				
			base:AddressBook				
			name	名字	Name	填写zh	
ATname-type	名字类型		ATname-type				
prefix	称谓		Prefix				
last-name	姓		LastName				
orgname	组织的完整名称		Orgname				
first-name	教名或姓名首字母		FirstName				
middle-name	中名		MiddleName				
suffix	名后缀		Suffix				
id	个人的标识号码		IndividualID				
role	职务		Role				
orgname	组织名称		Orgname				
department	部门名		Department				
synonym	同义名		SynonymicalName				
registered-number	注册号		RegisteredNumber				
addressbook	地址簿		Addressbook				

表中关于地

址信息的拆分及规范化加工参照了'中国区域代码字典':

中国基础数据 XML 样例如下：

```
<assignee>
  <assignee>
    <addressbook lang="zh">
      <name>海安常大技术转移中心有限公司</name>
      <address>
        <text>2226600 江苏省南通市海安县城东镇迎宾大道8号</text>
      </address>
    </addressbook>
  </assignee>
</assignee>
```

对应的中国标准化成品 XML 数据样例如下：

省份	市	区县	区域代码
北京市	市辖区	东城区	110101
北京市	市辖区	西城区	110102
北京市	市辖区	朝阳区	110105
北京市	市辖区	丰台区	110106
北京市	市辖区	石景山区	110107
北京市	市辖区	海淀区	110108
北京市	市辖区	门头沟区	110109
北京市	市辖区	房山区	110111
北京市	市辖区	通州区	110112
北京市	市辖区	顺义区	110113
北京市	市辖区	昌平区	110114
北京市	市辖区	大兴区	110115
北京市	市辖区	怀柔区	110116
北京市	市辖区	平谷区	110117
北京市	县	密云县	110228
北京市	县	延庆县	110229
天津市	市辖区	和平区	120101
天津市	市辖区	河东区	120102
天津市	市辖区	河西区	120103
天津市	市辖区	南开区	120104
天津市	市辖区	河北区	120105
天津市	市辖区	红桥区	120106
天津市	市辖区	东丽区	120110
天津市	市辖区	西青区	120111
天津市	市辖区	津南区	120112
天津市	市辖区	北辰区	120113
天津市	市辖区	武清区	120114
天津市	市辖区	宝坻区	120115
天津市	市辖区	滨海新区	120116
天津市	县	宁河县	120221
天津市	县	静海县	120223
天津市	县	蓟县	120225
河北省	石家庄市	长安区	130102
河北省	石家庄市	桥西区	130104
河北省	石家庄市	新华区	130105
河北省	石家庄市	井陉矿区	130107
河北省	石家庄市	裕华区	130108
河北省	石家庄市	藁城区	130109
河北省	石家庄市	鹿泉区	130110

cn-related-publication		相关公开文档	RelatedDocuments	加工规则（无需转换则取原值）	
cn-publication-reference	相关公开信息	相关公开信息	RelatedPublicationDoc		
lang	语言	语言	lang	转为小写字母	
country	国家	国家	country		
document-id	文献标识	文献标识	DocumentID		
ATlang	语言	语言	ATlang	转为小写字母	
country	国家	国家	WIPOST3Code		
doc-number	文献号	文献号	DocNumber		
kind	文献种类	名称	Kind		
name	名称	名字类型	Name		
ATname-type	名字类型	日期	ATnameType		
date	日期	专利信息/公报信息	Date		
gazette-reference	专利信息/公报信息	专利信息	GazetteReference		
id	ID标识	ATid	ATid		
country	国家	country	country		
lang	语种	lang	lang	转为小写字母	
gazette-num	公报号	GazetteNumber	GazetteNumber		
date	日期	Date	Date		
text	无结构文本	Text	Text		
on-related-documents		在先申请文档			
division	在先文档是该在 在先申请文档	Division	Division		
relation	文献关系	DocumentRelation	DocumentRelation		
parent-doc	父文献	ParentDocument	ParentDocument		

7、相关文档信息的整体结构和层级对应关系及加工规则如下：

```

<business:AssigneeDetails>
  <business:Assignee sequence="1" dateFormat="original" sourceBb="national office" lang="zh" creator="03">
    <base:AddressBook lang="zh">
      <base:Name>海安南大技术转移中心有限公司</base:Name>
      <base:AddressLine0></base:AddressLine0>
      <base:AddressLine1></base:AddressLine1>
      <base:AddressLine2></base:AddressLine2>
      <base:PostBox></base:PostBox>
      <base:PostRoom></base:PostRoom>
      <base:AddressFloor></base:AddressFloor>
      <base:AddressBuilding></base:AddressBuilding>
      <base:Street></base:Street>
      <base:AddressCity></base:AddressCity>
      <base:County>海安县</base:County>
      <base:City>南通市</base:City>
      <base:Province>江苏省</base:Province>
      <base:PostCode>226000</base:PostCode>
      <base:WIPOST3Code>CN</base:WIPOST3Code>
      <base:Text>226000 江苏省南通市海安县城东南镇太港8号</base:Text>
    </base:AddressBook>
  </business:Assignee>
</business:AssigneeDetails>

```

8、相关信息信息的整体结构和层级对应关系及加工规则如下:

[illegible]

相关人信息		Parties		加工规则（无需转换则取原值）		ApplicantDetails		见申请人代表性规则		Applicant		cn-applicant		代表人		AI cn-r		cn-applicants		申请人信息		AI sequence		顺序号		sequence		必须：有原值取原值，没原值按多值出现顺序排列 sequence=1， sequence=2 ...，同一语种使用同一套序号		creator		必须出现；两位小写对应字母，取zh		dataFormat		必须出现；取值original等		sourceDB		national 来源取值如 必须出现；数据 office等		ProcessingType		因多合一，初加工取值 original（表示原始） 申请人姓名在发明人中出现过，applicant-inventor，否则applicant		OrgType		参见申请人类型规则表		Base:AddressBook		AI lang		填写zh		Name		AI name-type		名字类型 natural, 自然人 legal, 法人实体 称谓		prefix		姓		OrgName		组织的完整名称		first-name		教名或姓名首字母		middle-name		中名		suffix		姓名后缀		id		个人的标识号码		role		职务		Role		OrganizationName		组织名称		Department		部门名		synonym		同义名		registered-number		注册号		RegisteredNumber		SynonymicalName		同义名		RegisteredNumber	
-------	--	---------	--	----------------	--	------------------	--	-----------	--	-----------	--	--------------	--	-----	--	---------	--	---------------	--	-------	--	-------------	--	-----	--	----------	--	--	--	---------	--	-------------------	--	------------	--	------------------	--	----------	--	---	--	----------------	--	--	--	---------	--	------------	--	------------------	--	---------	--	------	--	------	--	--------------	--	---	--	--------	--	---	--	---------	--	---------	--	------------	--	----------	--	-------------	--	----	--	--------	--	------	--	----	--	---------	--	------	--	----	--	------	--	------------------	--	------	--	------------	--	-----	--	---------	--	-----	--	-------------------	--	-----	--	------------------	--	-----------------	--	-----	--	------------------	--

[illegible]

[illegible]

相关信息-代理人的中国基础数据 XML 样例如下:

```
<business:ApplicantDetails representative="1">
  <business:Applicant sequence="1" type="applicant"
    dataFormat="original" sourceB="national office" lang="zh" creator="03"
    processingType="original">
    <base:AddressBook lang="zh">
      <base:Name>海安常大技术转移中心有限公司</base:Name>
      <base:Address>
        <base:AddressLine>0</base:AddressLine>
        <base:AddressMailCode>0</base:AddressMailCode>
        <base:PostBox>0</base:PostBox>
        <base:AddressRoom>0</base:AddressRoom>
        <base:AddressFloor>0</base:AddressFloor>
        <base:AddressBuilding>0</base:AddressBuilding>
        <base:Street>0</base:Street>
        <base:AddressCity>0</base:AddressCity>
      </base:Address>
      <base:Country>海安</base:Country>
      <base:City>南通市</base:City>
      <base:Province>江苏省</base:Province>
      <base:PostCode>226600</base:PostCode>
      <base:WIPOST3Code>CN</base:WIPOST3Code>
      <base:Text>226600
        江苏省南通市海安县城东镇迎宾大道8号</base:Text>
      </base:Text>
    </base:Address>
    </base:AddressBook>
    <business:OrganizationCode createdAt="00000000"
      creator="00">0000000000</business:OrganizationCode>
    </business:Applicant>
  </business:ApplicantDetails>
```

对应的中国标准化成品 XML 数据样例如下:

```
<cn-inventors>
  <cn-inventor sequence="1">
    <name>徐守坤</name>
    <cn-inventor request-by-inventor="no">
      <cn-inventor sequence="2">
        <name>石林</name>
      </cn-inventor>
    </cn-inventor>
  </cn-inventors>
```

相关信息-发明人的中国基础数据 XML 样例如下:

```
<business:ApplicantDetails representative="1">
  <business:Applicant sequence="1" type="applicant"
    dataFormat="original" sourceB="national office" lang="zh" creator="03"
    processingType="original">
    <base:AddressBook lang="zh">
      <base:Name>海安常大技术转移中心有限公司</base:Name>
      <base:Address>
        <base:AddressLine>0</base:AddressLine>
        <base:AddressMailCode>0</base:AddressMailCode>
        <base:PostBox>0</base:PostBox>
        <base:AddressRoom>0</base:AddressRoom>
        <base:AddressFloor>0</base:AddressFloor>
        <base:AddressBuilding>0</base:AddressBuilding>
        <base:Street>0</base:Street>
        <base:AddressCity>0</base:AddressCity>
      </base:Address>
      <base:Country>海安</base:Country>
      <base:City>南通市</base:City>
      <base:Province>江苏省</base:Province>
      <base:PostCode>226600</base:PostCode>
      <base:WIPOST3Code>CN</base:WIPOST3Code>
      <base:Text>226600
        江苏省南通市海安县城东镇迎宾大道8号</base:Text>
      </base:Text>
    </base:Address>
    </base:AddressBook>
    <business:OrganizationCode createdAt="00000000"
      creator="00">0000000000</business:OrganizationCode>
    </business:Applicant>
  </business:ApplicantDetails>
```

```
<business:Agency>
  <base:AddressBook lang="zh">
    <base:OrganizationName>
      北京众合诚成知识产权代理有限公司 11246</base:OrganizationName>
    </base:AddressBook>
  </business:Agency>
  <base:AddressBook lang="zh">
    <base:Agency>
      <base:AddressBook lang="zh">
        <base:Name>龚爱英</base:Name>
      </base:AddressBook>
    </business:Agent>
  </business:AgentDetails>

```

```
<business:AgentDetails>
  <business:CustomerNumber>11246</business:CustomerNumber>
  <business:Agent sequence="1" lang="zh" dataFormat="original">
    sourcedB="national office" repType="agent" processingType="original"
  </business:AgentDetails>
  creator="03">
```

对应的中国标准化成品XML数据样例如下:

```
<cn-agents>
<cn-agent>
<name>张焕英</name>
</cn-agent>
<cn-agents>
```

中国基础数据 XML 样例如下:

PCT 国际专利申请数据	document-id		文献标识	base: DocumentID	加工规则 (无需转换 则取原值)
	pct-or-regional-publishing-data		国际公布 有关数据	base: DocumentID	
	country		公开国家	WIPOST3Code	转为小写字母
	doc-number		公开(告) 号	DocNumber	
	kind		文献种类	Kind	
	name		名称	Name	
	属 name-type		名字类型 natural, 自然人	AT nameType	
	date		公开日期	Date	
	gazette-reference		专利信息/公 报信息	GazetteReference	
	id		id号	id	
	country		国家类型	country	转为小写字母
	lang		语言代码	lang	
	gazette-num		公告/公报编 号	GazetteReferen ce\GazetteMumb	
	date		公告/公报出 版日期	Date	
	text		公告/公报描 述性内容	base: Text	

对应的中国标准化成品 XML 数据样例如下:

```

<date-pct-article-22-39-fullfilled>
  <date>20151123</date>
</date-pct-article-22-39-fullfilled>
<date-pct-article-22-39-fullfilled>
  <country>JP</country>
  <doc-number>PCT/JP2014/065901</doc-number>
</date-pct-article-22-39-fullfilled>
<date-pct-article-22-39-fullfilled>
  <country>WO</country>
  <doc-number>2014/203859</doc-number>
</date-pct-article-22-39-fullfilled>
<date-pct-article-22-39-fullfilled>
  <country>JA</country>
  <doc-number>20141224</doc-number>
</date-pct-article-22-39-fullfilled>

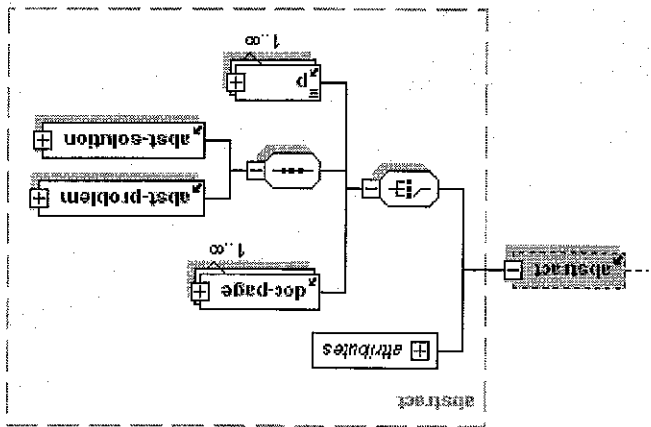
```

• DTD-Schema 结构如下:

2、中国标准化数据摘要结构

<abstract>
<p num="1">
本发明涉及一种机动车轮胎状态通讯检测平台,包括四个轮
胎数据采集仪和仪表盘控制器,所述四个轮胎数据采集仪分
别设置在所述机动车的四个轮胎里侧面对的车体壁面上,每
一个轮胎数据采集仪采集对应轮胎的状态信息,所述仪表盘
控制器与所述四个轮胎数据采集仪分别连接,基于接收到的
四个轮胎的状态信息,确定是否发出轮胎报警信号。通过本
发明,能够对轮胎的各个参数进行精确检测,提高轮胎报警
的准确性,保证机动车的安全行驶。</p>
</abstract>

• 样例数据结构如下:



• DTD-Schema 结构如下:

1、中国专利基础数据摘要结构

1.3摘要及摘要附图加工规则

```
<business:NationalStageEntryDate>  
<base:Date>20151123</base:Date>  
</business:NationalStageEntryDate>  
<business:RegionalApplicationData>  
<base:DocumentID>  
<base:DocNumber>PCT/JF2014/065901</base:DocNumber>  
<base:Date>20140616</base:Date>  
</base:DocumentID>  
</business:RegionalApplicationData>  
<business:RegionalPublishingData>  
<base:DocumentID lang="ja">  
<base:WIPOST3Code>W0</base:WIPOST3Code>  
<base:DocNumber>2014/203859</base:DocNumber>  
<base:Date>20141224</base:Date>  
</base:DocumentID>  
</business:RegionalPublishingData>
```

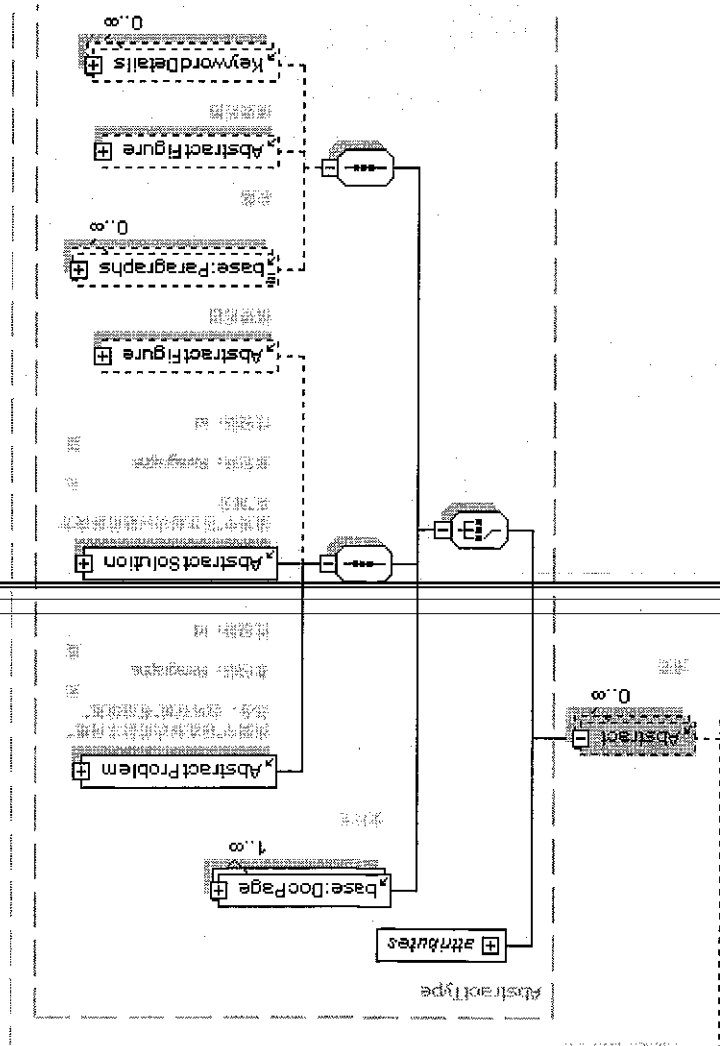
3、摘要及摘要附图标准化工规则如下：

```

<business:Abstract dataFormat="original" lang="zh" sourceDB
="national office" processingType="original" creator="03">
  <base:Paragraphs num="0001">
    本发明涉及一种机动车轮胎状态通讯检测平台，包括四个轮
    胎数据采集仪和仪表盘控制器，所述四个轮胎数据采集仪分
    别设置在所述机动车的四个轮胎里侧面对的壳体壁面上，每
    一个轮胎数据采集仪采集对应轮胎的状态信息，所述仪表盘
    控制器与所述四个轮胎数据采集仪分别连接，基于接收到的
    四个轮胎的状态信息，确定是否发出轮胎报警信号。通过本
    发明，能够对轮胎的各个参数进行精确检测，提高轮胎报警
    的准确性，保证机动车的安全行驶。</base:Paragraphs>
  <business:AbstractFigure>
    <base:Figure num="0001">
      <base:Image he="650.798071" wi="661" file="
      201510249243.TIF" imgFormat="TIF"/>
    </base:Figure>
  </business:AbstractFigure>
</business:Abstract>

```

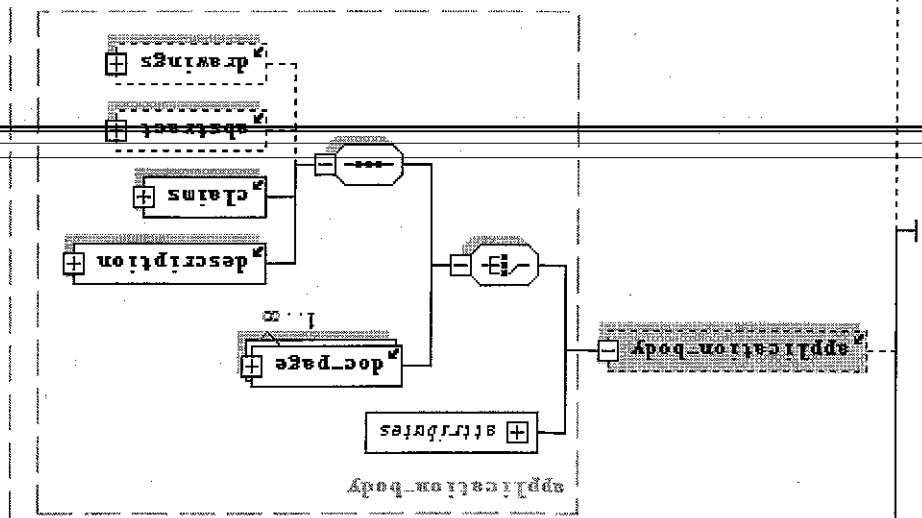
• 样例数据结构如下：



1.4 全文文本加工规则

[illegible]

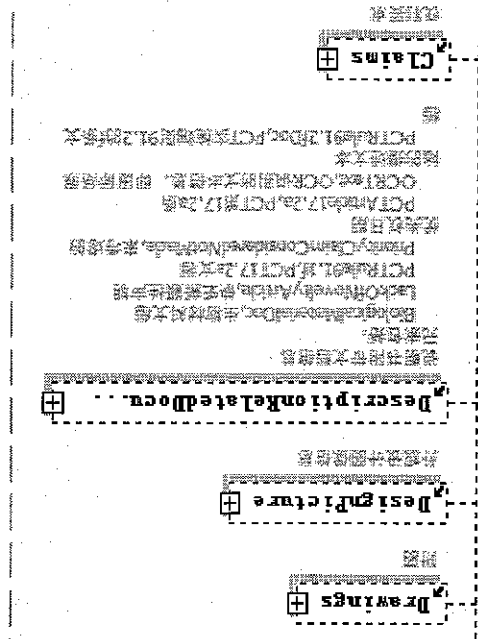
- DTD-Schema 结构如下:



- 样例数据结构如下:

2、中国标准化数据摘要结构

- DTD-Schema 结构如下:



● 样例数据结构如下:

1.4.1 说明书加工规则

Drawings	DesignPicture	DescriptionRelatedDocu	Claims
說明書	外觀圖	本發明係一種環保袋，其構造如下： 1. 袋體：由可降解材料製成，具有開口部及封口部。 2. 開口部：位於袋體之一端，用於裝填廢棄物。 3. 封口部：位於袋體之另一端，用於密封袋體，防止廢棄物洩漏。 4. 提手：連接於袋體之開口部，方便提攜。 5. 標籤：貼附於袋體表面，用於標示袋體之用途或品牌。	1. 一種環保袋，其構造如下： 2. 一種環保袋，其構造如下： 3. 一種環保袋，其構造如下： 4. 一種環保袋，其構造如下： 5. 一種環保袋，其構造如下：

[illegible]

[illegible]

[illegible]

对应的中国标准化成品XML数据样例如下:

```
<description>
<id="p0001" num="0001">技术领域
<id="p0002" num="0002">本发明涉及通讯检测领域,尤其涉及一种机动车轮胎状态通讯检测平
<id="p0003" num="0003">背景技术
<id="p0004" num="0004">机动车的出现方便了人们的出行,提供了人们生活的节奏和工作的效
<id="p0005" num="0005">现有技术中,存在一些机动车辅助驾驶机制,在机动车开动时,自动
<id="p0006" num="0006">因此,需要一种新的机动车轮胎状态通讯检测平台,能够根据采取多
<id="p0007" num="0007">发明内容
<id="p0008" num="0008">为了解决上述问题,本发明提供了一种机动车轮胎状态通讯检测平
<id="p0009" num="0009">根据本发明的一方面,提供了一种机动车轮胎状态通讯检测平台,所
<id="p0010" num="0010">更具体地,在所述机动车轮胎状态通讯检测平台中,还包括:四个高
<id="p0011" num="0011">更具体地,在所述机动车轮胎状态通讯检测平台中,所述无线收发器
<id="p0012" num="0012">更具体地,在所述机动车轮胎状态通讯检测平台中:所述液晶显示屏
<id="p0013" num="0013">更具体地,在所述机动车轮胎状态通讯检测平台中,还包括:双车道
<id="p0014" num="0014">更具体地,在所述机动车轮胎状态通讯检测平台中:每一个轮胎数据
<id="p0015" num="0015">附图说明
<id="p0016" num="0016">以下将结合附图对本发明的实施方案进行描述,其中:
<id="p0017" num="0017">图1为本发明实施例与现有技术中的机动车轮胎状态通讯检测平台的结
<id="p0018" num="0018">具体实施方式
<id="p0019" num="0019">下面将参照附图对本发明的机动车轮胎状态通讯检测平台的实施方
<id="p0020" num="0020">在过去的十年里,由于机动车性能大为改善,如安全带、安全气囊、
<id="p0021" num="0021">微软根大学发起了一项针对交通的新研究,对比了全球车祸死亡
<id="p0022" num="0022">为了提高机动车的行驶安全性,各个国家的机动车制造商纷纷为机动
<id="p0023" num="0023">为了克服上述不足,本发明搭建了一种机动车轮胎状态通讯检测平
<id="p0024" num="0024">图1为本发明实施例方案示出的机动车轮胎状态通讯检测平台的结
<id="p0025" num="0025">其中,仪表板控制器5与左侧轮胎数据采集仪1、右侧轮胎数据采集
<id="p0026" num="0026">接着,继续对本发明的机动车轮胎状态通讯检测平台的具体结构进行
<id="p0027" num="0027">所述检测平台还包括:四个高度测量仪,分别位于机动车四个轮胎所
<id="p0028" num="0028">所述检测平台还包括:用户输入器件,用于根据用户的输入,设定光
<id="p0029" num="0029">所述检测平台还包括:供电电源,包括太阳能电池器件、蓄电池、切
<id="p0030" num="0030">所述检测平台还包括:SDRAM,位于机动车仪表盘内,与所述用户
<id="p0031" num="0031">所述检测平台还包括:无线收发器,设置在机动车车体上,与所述仪
<id="p0032" num="0032">所述检测平台还包括:液晶显示屏,位于机动车仪表盘上,与所述仪
<id="p0033" num="0033">每一个轮胎数据采集仪与所述SDRAM连接,包括CMOS图像传感器、
<id="p0034" num="0034">所述图像识别器与所述CMOS图像传感器和所述SDRAM分别连接,
<id="p0035" num="0035">所述仪表盘控制器5与所述四个轮胎数据采集仪1-4、所述四个高
<id="p0036" num="0036">其中,所述图像预处理单元、所述轮胎检测单元和所述控制单元元
<id="p0037" num="0037">可选地,在所述机动车轮胎状态通讯检测平台中:所述无线收发器还
<id="p0038" num="0038">另外,PPGA(Field-Programmable Gate Array),即现场可编程门阵
<id="p0039" num="0039">以硬件描述语言(VHDL或Verilog)所完成的电路设计,可以经过简
<id="p0040" num="0040">PPGA一般来设计ASIC(专用集成电路)的速度要慢,实现同样的功
<id="p0041" num="0041">早在1980年代中期,PPGA已经在PLD设备中扎根.CPLD和PPGA
<id="p0042" num="0042">采用本发明的机动车轮胎状态通讯检测平台,针对现有技术中缺乏对
<id="p0043" num="0043">可以理解的是,虽然本发明已以较佳实施例披露如上,然而上述实施
</description>
```

<business:Description lang="zh" dataFormat="original" sourceB="national office" creator="03">

<business:InventionTitle id="title1">机动车轮胎状态通讯检测平台</

<base:Paragraphs id="p0001" num="0001">技术领域</base:Paragraphs>

<base:Paragraphs id="p0002" num="0002">

本发明涉及通讯检测领域,尤其涉及一种机动车轮胎状态通讯检测平台。</

base:Paragraphs>

<base:Paragraphs id="p0003" num="0003">背景技术</base:Paragraphs>

<base:Paragraphs id="p0004" num="0004">

机动车的出现方便了人们的出行,提供了人们生活的节奏和工作的效率,然而也带来了一些弊端,每年交通事故的死亡率已经与死于某些常见的健康问题的概率相当,如心脏病,这个数字是非常可观的,如何提高机动车的安全性,从客观上减少交通事故发生的数量,从而增加机动车市场占有率,是每一家机动车制造商都面临的重要课题。</

base:Paragraphs>

<base:Paragraphs id="p0005" num="0005">

现有技术中,存在一些机动车辅助驾驶机制,在机动车启动时,自动对机动车当前各种参数进行监控,以方便驾驶员及时了解机动车的当前状态,例如监控机动车周围障碍物的距离,监控机动车的加速度等。然而,由于轮胎是机动车与路面直接接触的唯一关键部件,现有技术中对机动车轮胎监控的方式仅仅限于胎压检测,即在机动车行驶过程中对轮胎气压进行实时自动监测,并对轮胎漏气和低气压进行报警,以确保行车安全,但这种检测方式仅基于单项参数的检测,可靠性不高,且仅限于胎压检测,无法检测轮胎胎体是否偏离位置。</base:Paragraphs>

<base:Paragraphs id="p0006" num="0006">

因此,需要一种新的机动车轮胎状态通讯检测平台,能够根据采取多项物理量同时检测的方式,综合对机动车轮胎状态进行判断,精确分析轮胎是否胎压不足以及是否偏移位置,从而有效避免交通事故的发生。</base:Paragraphs>

<base:Paragraphs id="p0007" num="0007">发明内容</base:Paragraphs>

<base:Paragraphs id="p0008" num="0008">

为了解决上述问题,本发明提供了一种机动车轮胎状态通讯检测平台,采用高度测量仪检测四个轮胎所在车轴位置距离地面的实时高度,采用图像检测装置检测四个轮胎的图像状态信息,基于每一个轮胎的实时高度和图像状态信息综合判断轮胎是否出现故障。</base:Paragraphs>

<base:Paragraphs id="p0009" num="0009">

根据本发明的一方面,提供了一种机动车轮胎状态通讯检测平台,所述检测平台包括四个轮胎数据采集仪和仪表盘控制器,所述四个轮胎数据采集仪分别设置在所述机动车的四个轮胎里侧面对的车体壁面上,每一个轮胎数据采集仪采集对应轮胎的状态信息,所述仪表盘控制器与所述四个轮胎数据采集仪分别连接,基于接收到的四个轮胎的状态信息,确定是否发出轮胎报警信号。</base:Paragraphs>

<base:Paragraphs id="p0010" num="0010">

具体地,在所述机动车轮胎状态通讯检测平台中,还包括:四个高度测量仪,分别位于机动车四个轮胎所在车轴位置,用于分别检测四个轮胎所在车轴位置距离地面的实时高度;用户输入器件,用于根据用户的输入,设定光线强度阈值、高度阈值、比例阈值、轮胎上限灰度阈值和轮胎下限灰度阈值,所述轮胎上限灰度阈值和轮胎下限灰度阈值用于将图像中的轮胎胎体与背景分离,所述轮胎上限灰度阈值和所述轮胎下限灰度阈值的取值都在0-255之间,所述轮胎上限灰度阈值大于所述轮胎下限灰度阈值;供电电源,包括太阳能

<base:Paragraphs id="p0042" num="0042">

采用本发明的机动车轮胎状态通讯检测平台,针对现有技术中缺乏对机动车轮胎全方位检测的有效方式的问题,采用图像采集处理技术和高度检测技术,实时获取四个轮胎所在车轴位置距离地面的实时高度和四个轮胎的图像状态信息,结合多种数据综合判断当前轮胎的状态,为驾驶员的驾驶提供更有价值的参考数据。</base:Paragraphs>

<base:Paragraphs id="p0043" num="0043">

可以理解的是,虽然本发明已以较佳实施例披露如上,然而上述实施例并非用以限定本发明。对于任何熟悉本领域的技术人员而言,在不脱离本发明技术方案范围情况下,都可利用上述揭示的技术内容对本发明技术方案做出许多可能的变动和修饰,或修改为等同变化的等效实施例。因此,凡是未脱离本发明技术方案的内容,依据本发明技术方案所以上实施例所做的任何简单修改、等同变化及修饰,均仍属于本发明技术方案保护的范围。

</base:Paragraphs>

</business:Description>

权利要求书数据部分整体结构和层级对应关系及加工规则如下:

1.4.3 附图信息加工规则

```

</business:Claims>
</business:Claim>
</business:ClaimText>
。其中，所述图像预处理单元还包括对比度增强处理子单元、中值滤波
子单元和灰度化处理子单元，所述中值滤波子单元采用3×3的像素滤波窗口执行中值滤波
</business:ClaimText>
其中，所述图像预处理单元还包括对比度增强处理子单元、中值滤波
子单元和灰度化处理子单元，所述中值滤波子单元采用3×3的像素滤波窗口执行中值滤波
芯片来实现；</business:ClaimText>
片的选型都为Xilinx公司的Artix-7系列，所述仪表盘控制器采用TI公司的TMS320DM642
芯片实现，所述图像预处理单元、所述胎体检测单元和所述控制运算单元所采用的FPGA芯
其中，所述图像预处理单元、所述胎体检测单元和所述控制运算单元分别采用不同的FPGA
</business:ClaimText>
供电器件以由所述太阳能供电器件供电，所述电压转换器与所述切换开关连接，以将通过
太阳能供电器件和所述蓄电池池剩余电量决定是否切换到所述太阳能
</business:ClaimText>
供电电源，包括太阳能供电器件、蓄电池、切换开关和电压转换器，所述切换开关与所述
255之间，所述轮胎上限灰度阈值大于所述轮胎下限灰度阈值；</business:ClaimText>
中的轮胎胎体与背景分离，所述轮胎上限灰度阈值和所述轮胎下限灰度阈值的取值都在0-
限灰度阈值和轮胎下限灰度阈值，所述轮胎上限灰度阈值和轮胎下限灰度阈值的取值都在0-
用户输入器件，用于根据用户的输入，设定光线强度阈值、高度阈值、比例阈值、轮胎上
</business:ClaimText>
轴位置距离地面的实时高度；</business:ClaimText>
四个高度测量仪，分别位于机动车四个轮胎所在车轴位置，用于分别检测四个轮胎所在车
</business:ClaimText>
</business:ClaimText>所述检测平台还包括：</business:ClaimText>
报警信号；</business:ClaimText>
述四个轮胎数据采集仪分别连接，基于接收到的四个轮胎的状态信息，确定是否发出轮胎
仪和仪表盘控制器，所述四个轮胎数据采集仪分别设置在所述机动车的四个轮胎里侧面对
的车体壁面上，每一个轮胎数据采集仪采集对应轮胎的状态信息，所述仪表盘控制器与所
1.一种机动车轮胎状态通讯检测平台，其特征在于，所述检测平台包括四个轮胎数据采
</business:ClaimText>
</business:Claim id="c10001" num="0001">
office creator="03">
</business:Claims lang="zh" dateFormat="original" sourceDB="national

```

对应的中国标准化成品XML数据样例如下:

```

</claims>
</claim>
</claim-text>其中，所述图像预处理单元还包括对比度增强处理子单元、中值滤波
</claim-text>其中，所述图像预处理单元、所述胎体检测单元和所述控制运算单元
</claim-text>所述仪表盘控制器与所述四个轮胎数据采集仪、所述四个高度测量仪
</claim-text>每一个轮胎数据采集仪与所述SDRAM连接，包括CMOS图像传感器、
</claim-text>液晶显示屏，位于机动车仪表盘上，与所述仪表盘控制器连接，用于
</claim-text>无线收发器，设置在机动车车体上，与所述仪表盘控制器连接，用于
</claim-text>SDRAM，位于机动车仪表盘内，与所述用户输入器件连接，用于接
</claim-text>供电电源，包括太阳能供电器件、蓄电池、切换开关和电压转换器，
</claim-text>用户输入器件，用于根据用户的输入，设定光线强度阈值、高度阈值、
</claim-text>四个高度测量仪，分别位于机动车四个轮胎所在车轴位置，用于分别
</claim-text>所述检测平台还包括：
</claim-text>1.一种机动车轮胎状态通讯检测平台，其特征在于，所述检测平台
</claim id="c10001" num="0001">

```

xsd/PatentDocument/Elements/FullImageElements.xsd

‘表 2.2-2-1’，所对应的.xsd 文件的相对路径为：/Appendix D

全文图像信息，主要是全文图像文件信息 FullDocImage，元素表详见

2 根据专利文献数据规范制定图像数据加工规则

对应的中国标准化成品 XML 数据样例如下:

对应的中国标准化成品XML数据样例如下：

中国基础数据XML样例如下:

[illegible]

BU0121004	ClaimPages	权利要求书图像页码范围
BU0121005	DescriptionPages	说明书图像页码范围
BU0121006	DrawingPages	说明书附图页码范围
BU0121007	DesignPicturePages	外观设计附图页码范围
BU0121008	OtherPages	其他附图页码范围

PDF 格式 (Portable Document Format)

所有该格式的文件应满足以下要求:

和 Adobe Acrobat v1.4 兼容; 没有压缩的文本; 没有加密的文本; 没

有数字签名; 所有的字体应是内嵌的并且是许可发布的; PDF 中图片的分

分辨率应在 150DPI 和 400DPI 之间, 推荐分辨率 300DPI; 整页图像尺寸采用 A4 (210×297mm) 格式, 其他图像的最大尺寸应小于 170mm×255mm。

2.1 图像页签加工规则

专利文献数据中, 全文图像文件作为单独文件, 推荐采用带标签的 PDF 文件格式, 并满足如下要求:

标签应每页设置; 标签名称格式: 三位大写字母+五位阿拉伯数字页

码信息; 标签名称由两级组成: 第一级三位大写字母。全文图像的各部分对应不同的三位大写字母缩写, 分别取其相应的英文说法的前三位字母。

如扉页 BIB、权利要求书 CLA、说明书 DES、说明书附图 DRA、其它 OTH; 第二级五位阿拉伯数字页码信息, 从 00001 开始, 按页码递增顺序编号;

其中, 扉页可有增补页。

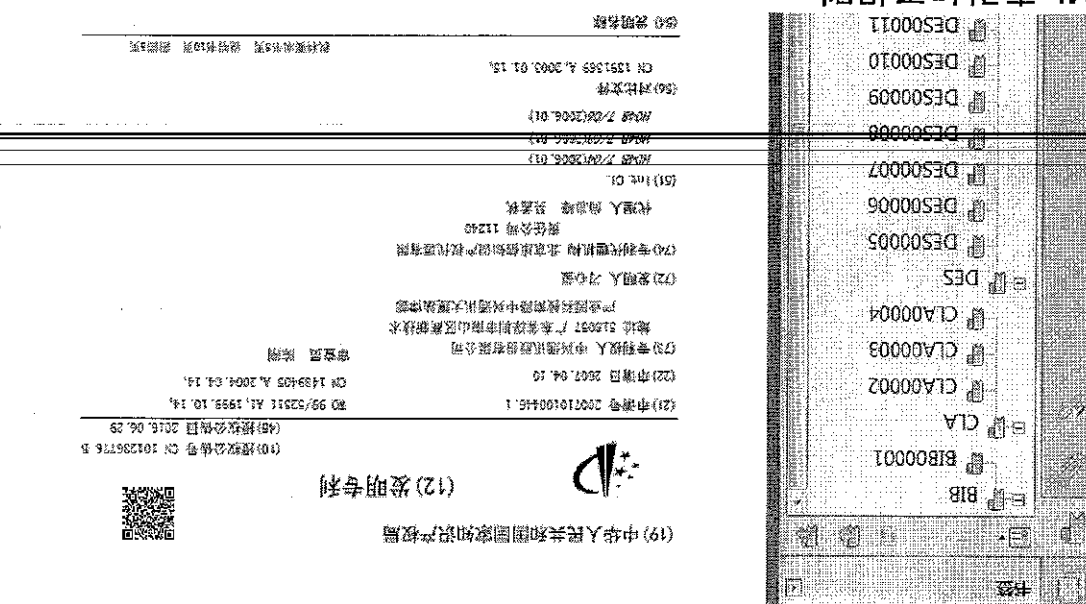
具体成品样例如下:

2.2XML索引加工规则

'ClaimPages','DescriptionPages','DrawingPages'等内容的索引内容，示例如

●

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<Business:PatentDocumentAndRelated xmlns:business="
20160622" datePublication="20160629" docNumber="CN101286776B" file=
"CN102007000100446CN00001012867760BPDFIDZH20160629CN00N.XML" kind="B" lang
="zh" status="C" xmlns:base="http://www.sipo.gov.cn/XMLSchema/base"
xmlns:m="http://www.w3.org/1998/Math/MathML" xmlns:tbl="
http://oasis-open.org/spec/soextblx" xmlns:xsi="
http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="
http://www.sipo.gov.cn/XMLSchema/business
/PTDS/PatentDocument/Elements/OtherElements.xsd" xsdVersion="V2.2.1">
<business:FullDocImage imagePath="
\CN102007000100446CN00001012867760BPDFIDZH20160629CN00N" type="pdf"
numberOfFigures="19" country="CN" lang="zh">
```



3 号码标准化转换规则

重要著录项目加工内容和要求

重要著录项目包括公开号、公开日、授权公告号、授权公告日、审定公告日、公告号、申请号、专利号、申请日、分类号和优先权。

表 2.2.3-1: 公开号

时间段	表示形式	备注	举例
在 1985-1988 年公	CN+申请号+A	无校验位, 仅发明	申请号: 85100134 公开号: CN85100134A
在 1989-2007.7.11	CN+7 位号+A	仅发明, 从 1030001 起	申请号: 89102485.9 公开号: CN1052186A
在 2007.7.18-至今	CN+9 位号+A	仅发明	申请号: 200610102290.6 公开号: CN100998279A

表 2.2.3-2: 授权公告号

时间段	表示形式	备注	举例
在 1993-2007.8.22	发明: CN+7 位号+C		申请号: 85101470

时间	表示形式	备注	举例
在 1985-至今 公开、授权的发明、实用新型专利	时间+国别代码+申请号	申请日、国别代码和申请号之间空半角空格	申请号：200310125451.X 优先权：2002.11.01 US 60/423335

表 2.2.3-6: 优先权

时间	表示形式	备注	举例
在 1985-1988 年公开的专利	XX X XXXXX 年 种类 号码	8 位, 无校验位	申请号：88210906 专利号：ZL 88210906
在 1989-2003 年 公开或授权的专利	XX X XXXXX.X 年 种类 号码 校验位	8 位, 有校验位	申请号：92224068.X 专利号：ZL 92224068.X
在 2003.10-至今 公开或授权的专利	XXXX X XXXXX.X 年 种类 号码 校验位	12 位, 有校验位	申请号：200310125451.X 专利号：ZL 200310125451.X

表 2.2.3-5: 申请号 (专利号)

时间	表示形式	备注	举例
在 1985-1988 年 公开的实用新型专利	CN+申请号+U	仅实用新型	申请号：88210906 公告号：CN88210906U
在 1989-1992 年 公开的实用新型专利	CN+7 位号+U	仅实用新型，从 2030001 起	申请号：92224068.X 公告号：CN125632U

表 2.2.3-4: 公告号

时间	表示形式	备注	举例
在 1985-1988 年 审定的发明专利	CN+申请号+B	仅发明	申请号：85100112 审定号：CN85100112B
在 1989-1992 年 审定的发明专利	CN+7 位号+B	仅发明，从 1003001 起	申请号：89100002.X 审定号：CN1014821B

表 2.2.3-3: 审定号

时间	表示形式	备注	举例
在 2007.8.29-至今 授权的发明、实用新型专利	发明：CN+9 位号+C 实用新型：CN+9 位号+Y		申请号：200310109960.3 授权公告号：CN100333631C 申请号：200620075737.0 授权公告号：CN200938735Y
授权的发明、实用新型专利	实用新型：CN+7 位号+Y		申请号：95230595.X 授权公告号：CN1023322C

公开日、授权公告日、审定公告日、申请日、优先权日等时间的表示方法统一为：yyyy.MM.DD 格式，其中 yyyy 表示年，MM 表示月，DD 表示日。例如：2016.01.01。

3.1 公开号标准化转换规则

遵照专利文献数据规范的定义，号码格式的形态由属性'dataFormat'控制，取值 standard、original、other 等，当 dataFormat='original'时，取值与加工前原始数据保持一致，当 dataFormat='standard'时，文献标识用

WIPOST3Code, DocNumber, Kind, Date 分别表示，分别遵循 WIPO 标准 ST.3, WIPO 标准 ST.6, WIPO 标准 ST.16, WIPO 标准 ST.9。公布号 DocNumber 应当仅由数字组成，数字总数量最大不超过 13，由各工业产权局根据需要确定，满足这些需求的数字位应尽可能短。由于原始数据的格式满足此要求，故定义中国标准化数据标准公开号与原数据公开号一致。

3.2 申请号标准化转换规则

遵照专利文献数据规范的定义，号码格式的形态由属性'dataFormat'控制，取值 standard、original、other 等，当 dataFormat='original'时，取值与加工前原始数据保持一致，当 dataFormat='standard'时，标准申请号需遵循 WIPO ST.3 和 ST.13，其中为 DocNumber 设计了 15 位字符的长度，即 2 位知识产权类型代码+4 位年代标识+9 位序列号。若无年代标识信息则 4 位年代标识取值为 0000。

2 位知识产权代码的规定同 WIPO ST.13:

- 10:用于发明专利申请
- 11:用于处于国家阶段的 PCT 申请
- 12-19:各工业产权局自行规定
- 20: 实用新型的申请
- 21: PCT 申请中的实用新型的申请
- 22-29: 各局专用

30-89: 各局专用

90-99: 为 WIPO 国际局的使用预留

91: PCT 国际阶段下的国际申请

原始申请号样式示例: YYYSSNNNN.N.X

```
<application-reference appl-type="发明专利">
  <document-id>
    <country>CN</country>
    <doc-number>201610249243.3</doc-number>
  </document-id>
</application-reference>
```

标准化申请号样式示例: SSYYYNNNNNNNN (N: 9 位左补零流水号)

```
<business:ApplicationReference applType="10" dataFormat="standard" sequence="1">
  <base:DocumentID>
    <base:WIPOST3Code>CN</base:WIPOST3Code>
    <base:docNumber>100015000249243</base:docNumber>
  </base:DocumentID>
  <base:Date>20160515</base:Date>
</business:ApplicationReference>
```

3.3 优先权号标准化转换规则

优先权号的标准化规范参照上一节申请号的规范,不同的是申请号是指中国专利申请号,而优先权号可以是世界各个国家的申请号。每个国家都有各自的申请号结构标准,而各国的申请号标准并不固定,可能根据各国专利法律的修改而变更格式,所以优先权号的标准化规则就需要按照各个国家、时间范围、号码长度、类型等特点具体制定。

以优先权为中国和英国的专利号为例,具体号码规则如下:

中国 (CN)					
位长	内容	条件	字母公式	样例	
14位长	4位年+8位流水号 +“+”1位校验位 3位字母+“/”+2位 字母+2位年+“/”+5 位流水	第13位为字 母,第4/8位为 第13位为字 母,第4/8位为 母,第4/11位为	PCT/CNYY/NNNNN PCT/CN04/00077	201110043647.9	
10位长	2位年+6位流水号	第9位为“+”,其 不包含“+”	YYNNNNN.X	99258987.5	
9位长	2位年+6位流水号	不包含“+”	YYNNNNNXX	94228271X	
17位长	3位字母+“/”+2位 字母+4位年+“/”+8 位流水	第13位为字 母,第4/11位为	PCT/CNYYYY/NNNNNN PCT/CN2010/078548		

型、号码年代、申请流水号、文献流水号、文献种类代码、数据资源代
专利文献数据文件名标识码通过提取每条数据记录的国别、专利类
E.1 所示。

专利文献数据文件名标识码由 52 位数字/字母/符号组成，其结构如图
图 E.1 为专利文献数据文件名标识码的结构

文件名标识码的结构及各组成项需要按照以下规范命名。
依据已制定的《专利文献数据规范》(ZC 0014—2012)，专利文献数据
1、标准化数据的文件名命名规则

4 中国标准化数据文件名及层级结构规则

位长	内容	条件	字母公式	样例	备注
9	2位年+5位流水号 +1位校验 第8位是“ ” 其余位是纯数字	YYNNNN.X 0301117.8			
	2位字母+2位年+5位流水号 前2位是字母，其余位是纯数字 流水号	1LYNNNN GB8607399			
	2位年+5位流水号 2位年+5位流水号 流水号 +1位校验 第9位是“ ” 其余位是纯数字	YYNNNN 85.05811		第3位还可换，或-或/代替	
8	5位流水号 +/+2位年	NNNN/YY 18397/86			
	2位年+5位流水号 2位年+5位流水号 流水号 +1位校验 第9位是“ ” 其余位是纯数字	YYNNNN.X 02250852.7			
10	2位年+5位流水号 2位年+5位流水号 流水号 +1位校验 第9位是“ ” 其余位是纯数字	YYNNNN.X 90.07970.8			
	2位字母+2位年+5位流水号 前2位是字母，第10位是“ ”，其余位是纯数字	1LYNNNN.X GB9406618.0			
11	2位年+6位流水号 +1位校验 第10位是“ ”，第2位是“/”，第10位是“ ”	YY.NNNNN.X 89/015736.8			

英国 (GB)

码、语言代码、法律生效日期、数据来源标识、备用位、校验位等十三项信息，实现对每条专利数据的唯一标识。

专利 类型	年份	申请流水号	国别	文献 流水号	种类	数据资源	语言	法律生效	来源	备用位	校验
C 1	K 1 YY YY	NNNNNNNNN	C 2 NNNNNNNNNN	K 2 NNNNNNNNNN	SR C	L L YY YY M M D D S S	YY YY MM DD	YY YY MM DD	YY YY MM DD	YY YY MM DD	YY YY MM DD
2位 代码	2位 专利 种类	4位 年份	9位 顺序 号,不足 的左补0	2位 国别 代码	13位 顺序 号,不足 的左补0	2位 种类 字符	2位 语言 代码	YY YY MM DD 格式的日期	4位 数据 来源	5位 备用 号	5位 校验 码
3	4	5	6	7	8	9	10	11	12	13	14

图 E.1 专利文献数据文件名称标识结构图

E.2 为专利文献数据文件名称标识码各组成项的规范

专利文献数据文件名称标识码的十三个组成项的内容、规范和说明如下:

E.2.1 国别/地区/组织代码 (C1) (第1-2位)

标识申请数据的国别/地区/组织代码信息, 2位, 取值参照 WIPO ST.3 标准。

对于无法确定国别/地区/组织代码信息的数据, 相应位置补 00。

E.2.2 专利类型 (K1) (第3-4位)

——标识专利类型的代码, 2位, 取值参照 WIPO ST.13 标准。

——特殊专利数据类型代码

对于 ST.13 标准未包含的专利数据, 3-17 位从右至左存放原始号码信息, 不足位数的左补零。

——非专利数据类型代码

用8符号代替 DOI 号码各部分的分隔符; 去掉 DOI 号码的第4部分的0和第5部分的“”符号, 其余保留原格式, 右对齐, 存放于3-32位, 不足位数的左补零; DOI 号码的格式见表 E.1。

——专利类型无法获知的情况下, 赋值 00。

表 E.1 DOI 号码的源格式

DOI 号码的部分	最大位数	内容	备注
一	4	刊物来源	CNKI : SCN SUN ISSN
二	1	刊物类型	
三	4	刊物代号	
四	1	以“-”分隔刊物年份、期号及引文起始页码，	最大位数的计算中包含分隔位数，共计 33 位
五	11		
示例			
			1671-6205
			ZGCF
			11-3014
			2007-01-001
			2008-03-001
			1999-06-001

表 E.2 DOI 号码的标准格式

DOI 号码的部分	最大位数	内容	备注
一	4	刊物来源	CNKI & ISSN & SUN & SCN
二	1	刊物类型	
三	4	刊物代号	
四	1	以“-”分隔刊物年份、期号及引文起始页码，	最大位数的计算中包含分隔位数，共计 33 位
五	11		
示例			
			1671-6205
			ZGCF
			11-3014
			200701001
			200803001
			199906001

E.2.3 年代信息 (yyyy) (第 5-8 位)

- 年代信息，4 位，不足的左补齐 18，19 或 20 等。
- 对于号码中无年代信息的，从申请日期中提取年代信息；
- 采用特殊纪年的号码，应将其年代变换为公元纪年。
- 无法获知的情况下，赋值 0000。

表 E.2 数据资源名称及代码

E.2.4 申请流水号(NNNNNNNN) (第 9-17 位)	申请流水号信息, 取值参照 WIPO ST.13 标准, 赋值 9 位, 不足位数的左补 0。 E.2.5 国别/地区/组织代码 (C2) (第 18-19 位) 标识文献数据的国别/地区/组织代码信息, 赋值 2 位, 取值参照 WIPO ST.3 标准。
E.2.6 文献流水号(NNNNNNNNNNNN) (第 20-32 位)	文献流水号信息, 取值参照 WIPO ST.6 标准, 赋值 13 位, 不足位数的左补 0。 E.2.7 文献种类代码(K2) (第 33-34 位) ——以文献为数据单位的数据记录, 参照 WIPO ST.16 标准, 赋值 2 位, 不足的左补 0; ——以专利族为数据单位的数据记录。 ——其他特殊数据的文献种类代码赋值: 上述约定未包含的其他情况, 根据实际业务需要对文献种类代码的赋值进行补充。但是, 新补充的取值不应与上述取值冲突, 并应对新取值的含义进行说明。
E.2.8 数据资源代码(SRC) (第 35-37 位)	标识专利文献数据记录的数据内容的代码, 可为数字或字母, 赋值 3 位, 具体如表 E.2 所示;

标识数据所采用语言的代码，赋值 2 位。

——仅对应一种语言的专利文献数据，语言代码取值参照 ISO639-1，两位小写字母。

——同时存在多种语言的专利文献数据，语言代码取值为专利文献数据文件中语言种类的数量，不足两位则左补零；语言种类数量无法获取的情况下赋值 00。

E.2.9 语言代码(LL) (第 38-39 位)

含义进行说明。

值进行补充。但是，新补充的取值不应与上述取值冲突，并应对新取值的上述约定未包含的其他情况，应根据实际业务需要对数据资源代码赋

信息	Modules/ChemicalSubstanceInformationCHS.xsd		
中药	方剂	TCM	Modules/MedicalFormulaInformationTCM.xsd
标题	TIT	Modules/InventionTitleTIT.xsd	
国际	专利	ICR	Modules/ClassificationPCRDetailsICR.xsd
其它	分类	ICO	Modules/OtherClassificationDetailsICO.xsd
机构	OCO	Modules/PatentDocument/	Modules/OrganizationCodeOCO.xsd
代码			
复审	决定	DEC	Modules/AppealDecision/
司法	判决	DIC	Modules/PatentVerdict/
法律	状态	PRS	Modules/PRS.xsd
引文- 引用	引文- 信息	CIT	Modules/CitingAndCitedDocuments/
引文- 被引	信息	CIG	Modules/CitingAndCitedDocuments/
专利	族	FAM	Modules/PatentFamily/
.....			Modules/PatentFamilyFAM.xsd

专利文献数据 文件名标识码	CN	中国别代码
示例数据 52 位 文件名信息		注释

左到右顺序对应如下表:

按照专利文献数据文件名标识码的命名规范, 示例中 52 位文件名从

 CN102014000293651CN00001056152600APIDZH20160601CN005
 CN102014000293651CN00001056152600APDFZH20160601CN004.PDF
 CN102014000293651CN00001056152600APIDZH20160601CN005.XML

(2) 图像示例:

 CN102015000249243CN000010484272008FULZH20160106CN00R
 CN102015000249243CN000010484272008FULZH20160106CN00R.XML

(1) 文本示例:

专利文献数据文件名标识码的示例图如下:

示例:

采取 ISO 7064 MOD 37-2 校验码计算法。

专利文献文件名标识码的校验位, 赋值 1 位数字或字母或附加符*,

E.2.13 校验位(X) (第 52 位)

——无需启用备用位时, 取值 00。

用位, 赋值 2 位, 根据实际业务需要对赋值取值进行补充;

——以上标识信息仍无法唯一标识专利文献数据的情况下, 须启用备

E.2.12 备用位(BK) (第 50-51 位)

标识该数据的来源的信息, 赋值 2 位。

E.2.11 数据来源标识(SS) (第 48-49 位)

识作用的情况下, 赋值 00000000。

日期无法获知、日期缺乏存在的意义、以及无须日期信息即可起到标

——其他特殊数据的法律生效日期赋值:

——标识数据资源生效的日期, 采用 YYYYMMDD 格式, 赋值 8 位。

E.2.10 法律生效日期(YYYYMMDD) (第 40-47 位)

专利类型	10	发明专利
号码年代	2015	2015 年
申请流水号	000249243	
国别	CN	中国国别代码
文献流水号	0000104842720	
种类代码	0B	
数据资源代码	FUL	《表 E.2 数据资源名称及代码》中无 FUL
语言代码	ZH	中文
法律生效日期	20160106	授权公告日
来源	CN	中国
备用位	00	
校验	R	

2、标准化文本数据的层级结构规则

中国专利标准化数据的层级结构按照以下方式：

数据类型\公开日\流水号\52 位文件夹名\数据文件

文本数据示例：

full text for inventiongrant

\20160106\4\CN102015000249243CN000001048427200BFULZH20160106

CN00R\

图像数据示例：

full image for invention grant

\20160601\1\CN102006000172022CN00000019875240BPIDZH20160601

CN00S\

各层级的具体结构内容规则如下：

数据类型有以下三种：

——full text for invention grant (发明授权)

——full text for invention publication (发明公布)

——full text for utility model (实用新型)

公开日：

公开日按 8 位日期格式。

流水号

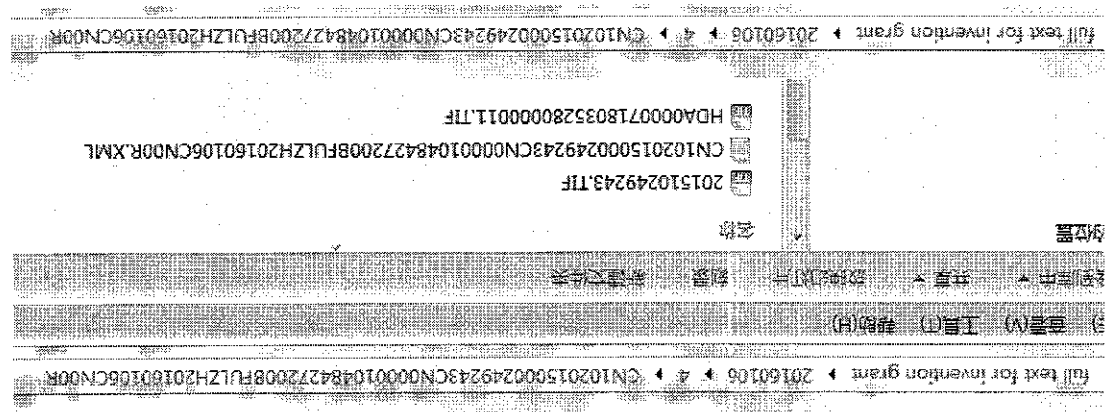
流水号层级是每 2000 个单位为一个文件夹，以流水号 1, 2, 3...命名文件夹并往下排列。

52 位文件夹名

一一命名依据同专利文献数据文件名标识码的命名规范。
数据文件

包含以 52 位文件名命名的 XML 格式数据及图像数据。

层级结构在 Windows 资源管理器中的示例展示图如下：



中国专利数据标准化加工服务项目质检规则

针对本项目，知识产权出版社有限责任公司会遵守质量检测体系的标准要求，制定具体、合理、可行的项目质量检测措施，确保成果符合招标文件及招标人需求，即确保质量检测体系中包含招标方要求的对加工完成的成品数据进行机器全量质检、对每次提交的成品数据进行不低于10%的人工抽检和对质检不合格的数据批次返回重新加工的需求。

5 质量检测目标

- 1、确保交付的数据加工服务方案总体架构、数据加工方案、实施计划方案、系统设计、质检报告及其他应交付的文档质量达到行业标准；
- 2、确保交付的全部设计内容经过内部审查和评审，具有完整性、一致性和方便性；
- 3、保证最后成果符合招标文件及招标人需求，最终实现客户满意。

6 质量检测原则

- 1、以用户需求、验收标准为质量检测的依据，采用合理质量检测措施。
- 2、项目质量检测措施贯穿项目实施全过程。

7 质量检测方式

根据检查内容的不同，分别采用机器全量质检和按比例人工抽检两种方式。其中，数据形式检查主要以机器全量质检为主，人工抽检为辅；数据的实质内容检查主要以人工抽检为主，机器质检为辅。

1、机器全量质检

通过对待质检成品数据进行前期调研，收集、汇总、整理和分析相关资料或信息（主要包括加工规则或标准、质量要求或标准等），明确数据加工可能存在的风险，找准质量监测点及监测标准，完成机器全量质检方案。按照质检方案进行质检工作。包括对提交数据文件的规范性、数据内容的完整性、数据格式正确性等内容进行程序辅助的机器全量检查，并填写质检记录单。

2、人工按比例抽检

通过对待质检成品数据进行前期调研，收集、汇总、整理和分析相关资料或信息（主要包括加工规则或标准、质量要求或标准等），明确数据加工可能存在的风险，找准质量监测点及监测标准，完成人工按比例抽检方案。按照质检方案进行质检工作。检查为按照数据加工规则或标准以及质量要求或标准，对于数据内容的正确性进行检查，并填写质检记录单。对数据的实质检查采用人工抽样的方式进行。

8 内部质量检测流程

质量控制总体流程，如下图所示：

合分析，确定关键质量控制点的质量控制标准，编写相应的质量保障工作

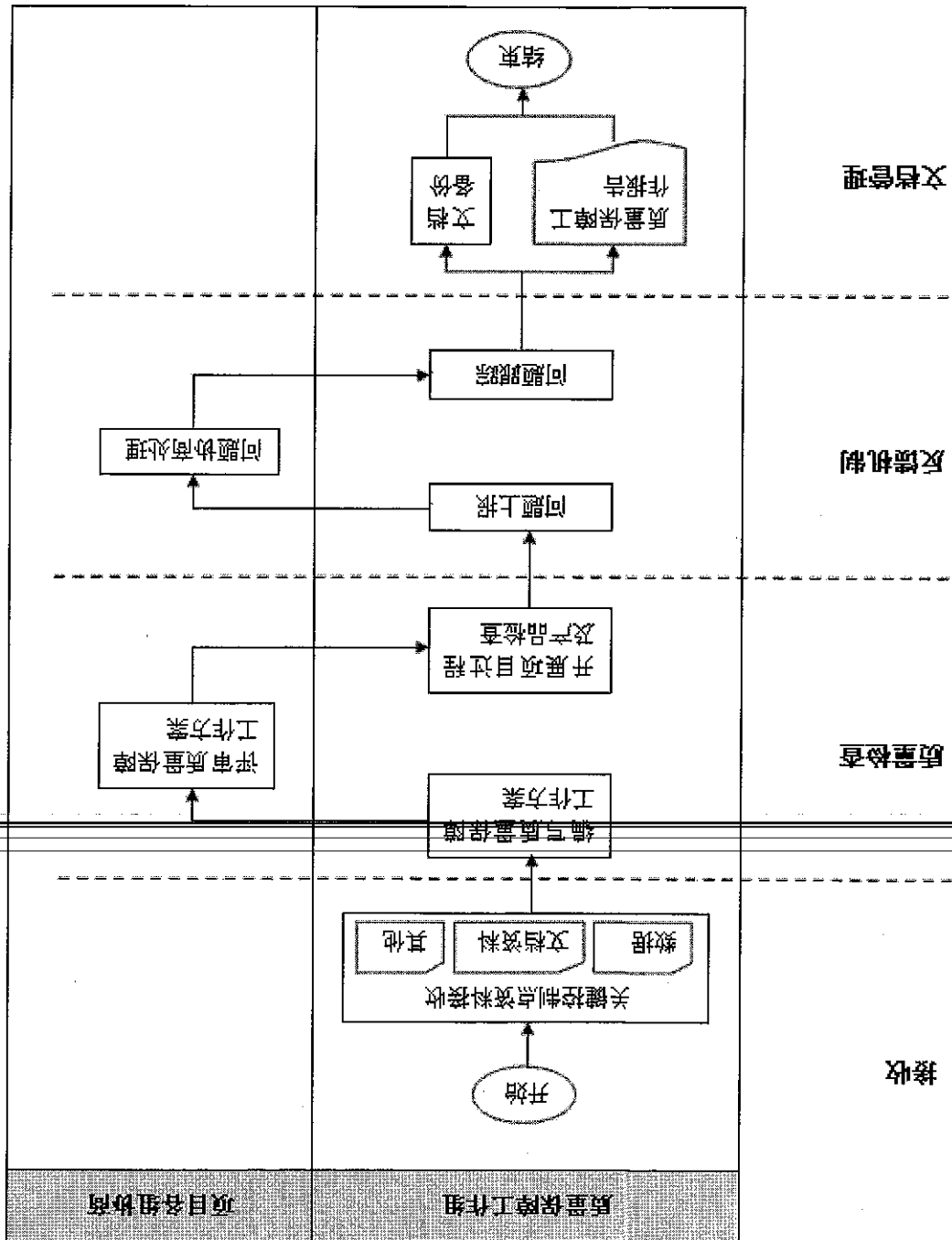
2、质量检查——根据在关键质量控制点收集到的各类材料，进行综

收、汇总和整理。

量控制点，在关键质量控制点进行源数据、成品数据及相关文档资料的接

1、接收——对整个数据项目的全过程进行分析、分解，确定关键质

质量控制流程说明如下：



方案，并与相关部门一并进行质量保障工作方案的评估，进行反复的修改完善，待评估通过后，即可进行项目全过程的质量控制及产品检查。

3、反馈机制——对于关键质量控制点监控发现的质量问题，要及时

反馈到各个相关环节，及时采取修正措施，进行问题跟踪，直至问题解决，从而形成了质量管理闭环的监控机制。反馈机制包含两个方面：

- 对外反馈机制——质量保障工作组通过发现的问题及梳理出的问题

易发点，来提出合理可行的改进建议反馈给数据加工工作组及管

理组，指导并完善数据资源加工、质检及应用服务体系，从源头实施对于数据资源的监控，切实提高数据资源的质量，以满足客户的质量需求。

- 对内反馈机制——质量保障体系运作过程中，对于质量保障工作组内部产生的各种工作流程或工作机制方面的问题也应及时进行总结，并进行相应调整，促进质量控制流程的顺利进行及质量保障体系的高效运作。

- 文档管理——流程进行到后端，经质量控制评估合格的数据，质量保障工作组生成质量保障工作报告，提交相关部门及领导；在整个项目执行过程中，形成的各类管理文档、质量保障方案、质量保障工作报告等应分门别类地进行归档备份，有助于资料的长期留存、有效信息的随时查阅，当项目人员存在变动时，还有助于变动人员及时了解项目情况，保证了信息的可持续性使用。

按照数据资源加工的质量要求，对于加工数据的质量有两种监督方

式：（1）项目组内部派遣专业的质检人员对加工后的成品数据进行内部的质检和验收。（2）为了避免项目质检可能存在的质检盲点，在项目组之外设立独立的第三方质检小组，负责数据资源的第三方抽检工作。

8.1 项目组内部质检工作流程

成品数据质量检查流程如下图所示：

通过对待质检成品数据进行前期调研，收集、汇总、整理和分析相关资料或信息（主要包括加工规则或标准、质量要求或标准等），明确数据加工可能存在的风险，找准质量监测点及监测标准，完成质检方案。按照质检方案进行质检工作。

数据的形式检查包括对提交数据文件的规范性、数据内容的完整性、数据格式正确性等内容进行检查，并填写质检记录单。数据的形式检查根据情况可以进行全检，也可进行抽检。

数据的实质性检查为按照数据加工规则或标准以及质量要求或标准，对于数据内容的正确性进行检查，并填写质检记录单。对数据的实质检查采用程序抽样的方式进行。

4、检查方式

根据检查内容的不同，分别采用机器全量质检和按比例人工抽检两种方式。其中，数据形式检查主要以机器全量质检为主，人工抽检为辅；数据的实质检查主要以人工抽检为主，机器质检为辅。

5、质检结果的处理

数据质检完成后，对数据质检结果进行整理、分析、反馈、复检，并撰写完成质检报告。问题数据如经复检无问题，数据质检结果达到数据质量控制标准的数据为合格，允许进入下一流程，如未能达到数据质量控制标准的数据，须返工重新加工后再提交。

6、质量检查报告

导出检查记录主要是指在执行质检任务中产生的统计表格和检查结果。包括导出质检记录单、导出质检报告。最终核实无误的质检报告可推送给相关领导审阅。

此外，最终的质检方案、质检记录单和质检报告可按要求进行存储、备份和集中管理，以便日后查阅。

下表是加工数据内部检测报告：

数据检测报告		送检批次（时间）	文献范围	批量（记录数）	样本量（记录数）	检测项目	错误数	校验员	检测完成时间	检验负责人	检测部门	部门审核人
备注												

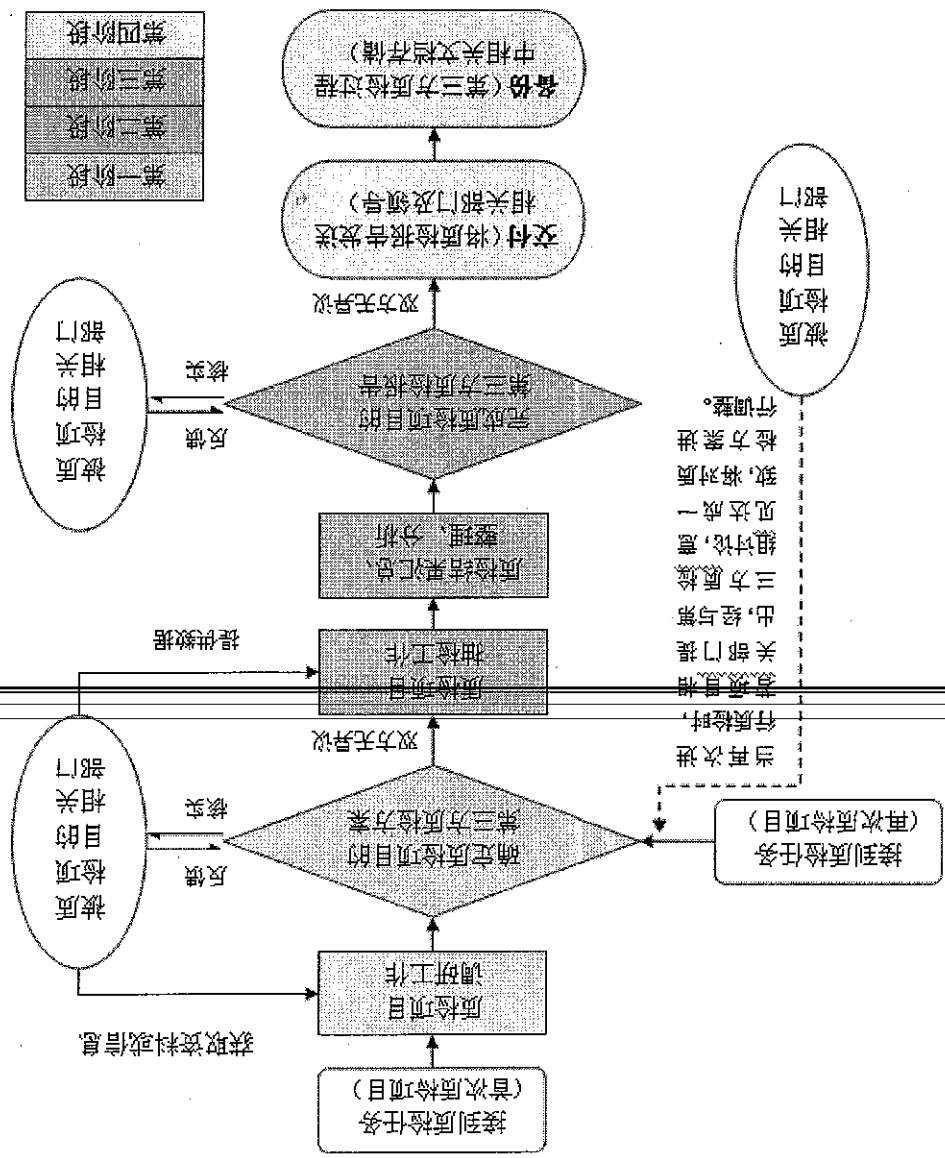
8.2 项目组外部第三方质检工作流程

第三方质检是在项目组外设立的、独立于数据资源加工、处理及项目质检流程之外的一种数据质量监督和检测体系。第三方质检工作对数据质量进行客观评价，并针对第三方质检结果进行深入调研和分析，从数据加工和质检规则、流程以及人员和软硬件配置等诸多方面向项目组提出改进的意见和建议，从源头实施对于数据资源的监督与控制，保障数据资源的质量。

第三方质检工作流程如下图所示：

- 若该项目为首次质检时，需要对质检的项目进行项目调研，主要包括收集、汇总、整理和分析该项目的相关资料或信息，撰写完成针对该项目的第三方质检方案。该质检方案需与该项目组进行核实、反馈意见，直至第三方质检组与该项目组意见达成一致，并对该项目的第三方质检方案无异议，方可确定该质检项目的第三方质检方案。
- 若该项目为再次质检时，如该项目组对其质检规则、加工规则等指标进行调整和变化时，经与第三方质检组讨论，达成一致意见

1、第一阶段确定质检方案



后，第三方质检组需要对之前制订的第三方质检方案进行相应修改和调整，并与该项目组确认，双方无异议，方可确定该质检项目的第三方质检方案。同时收集该项目调整修订后的有关档案资料。

上述项目相关资料包括：

A.项目简介,主要包括项目背景、数据来源、数据范围、加工方法（包括机器加工、人工加工等）；

B.数据加工项目的加工相关文档，主要包括：加工规则或加工标准、加工内容及工艺流程、加工流程或工具（软件）、加工进度及更新情况；

C.数据加工项目的内部质检相关文档，主要包括：质检规则或质检标准、质检内容或范围、质检流程或工具（软件）、质检方式（包括机器质检、人工质检、抽检、全检等）、质检要求的合格率或合格标准；

D.数据加工项目的内部质检记录单或质检报告；

E.其他，根据项目的特殊性需要特别说明的内容。

2、第二阶段数据质检

第一阶段工作完成后直接进入第三方质检的第二阶段，对该项目组提供的成品数据，按照确定的质检方案进行抽样质检工作，填写抽检记录单。

根据项目情况，有些项目的质检工作将分配给质检辅助人员完成数据质检工作。

3、第三阶段完成质检报告

数据质检完成后，进入第三方质检的第三阶段，对数据质检结果进行整理、分析，并撰写完成该项目的第三方质检报告。

该质检报告需与该项目相关部门核实，反馈意见，直至第三方质检组
与该项目相关部门意见达成一致，并对该项目的第三方质检报告无异议，
方可确认该质检项目的第三方质检报告完成。

4、第四阶段质检报告的交付与备份

最终确认无误的第三方质检报告将发送给该项目组领导审阅。

同时，第三方质检组将对质检方案、抽检记录单、质检报告，以及上
述第三方质检过程中的相关文档（如加工规则、质检标准等）进行存储、
备份和集中管理。

下表是加工数据第三方质检报告：

数据质检报告		
数据范围		
批量（记录数）		
	内部质检	第三方质检
送检时间		
样本量（记录数）		
抽检比例		
检测项目		
差错率（万分率）		
结论		
质检员		
检测完成时间		
检测部门		

9 内部质量检测环节

9.1 标准化文本数据成品检测

流程图如下：

针对日期

正则表达式: ^[A-Z]{2}\$

规则: 两位大写国别字母 CN

针对国别

尾。

\$表示以它前面的一个字符结尾, 这里就是以[a-z]范围内的字母结尾。
{2}, 是指前一个字符的数量, 这里表示两个[a-z]范围内的字母。
[], 是范围选择的意思。这里[a-z]表示字符选择的范围是从 a 到 z。
示以 a 到 z 的小写字母开始。

“^”, 正则表达式第一位的“^”表示以后面字符开始, 在这个正则中表
正则释义:

正则表达式: ^[a-z]{2}\$

规则: 2 位小写国别字母 zh

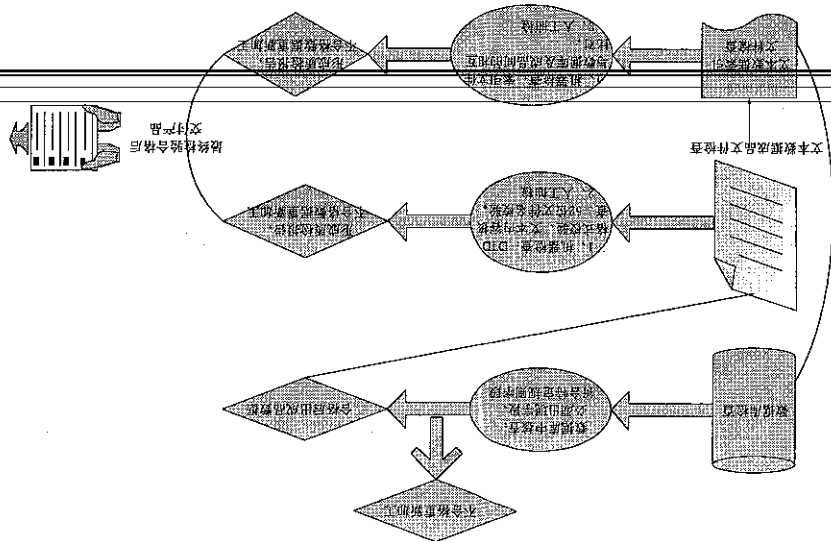
针对 lang 属性

质检规则

通过数据库环节的检查, 在前端发现数据问题, 及时反馈修正。1、

第一步: 数据库检查:

该成品检测分三步:



规则：8 位数字

正则表达式：\d{8}\$

正则释义：\d'表示数字，也可以用范围的方式表示数字'10-9'。这里写成'\[0-9]{8}\$'是同样的意思

针对 dataformat

规则：standard 在数据库中对应的值是 1,original 在数据库中对应的值

是 2。

针对 sourceDB

规则：dataFormat='original'时，sourceDB='national office'必须出现；

表与表之间的大致关系

在中国的数据库中表 T_EXCHDOC 是根表，相当于标准化之后 XML 文档的根节点。T_EXCHDOC 的 RID 是其他各个表的 P_RID(引证部分例外)。因此只要知道一件专利的公开号和公开类型就可以查出来大部分想要的节点。

2、具体表格信息

T_ABS_ABSTRACT

需要校验的字段：

ATT_LANG	适用于 lang 语言的规则。参考第一部分 a。
----------	--------------------------

ATT_DATAFORMAT	适用于 dataformat 的规则，参考第一部分 d。
----------------	------------------------------

ATT_DATASOURCE	适用于 datasource 的规则，参考第一部分 e。
----------------	------------------------------

ATT_DATE	适用于日期的规则，参考第一部分 c。
----------	--------------------

ATT_COUNTRY	两位大写国别字母，参考第一部分 b。
-------------	--------------------

ATT_PROTYPE		
原始XML中的值	数据库中对应的值(查询条 件)	备注
original	1	发明创造标题加工方法:原始

T_EXCHDOC

COUNTRY		
两位大写国别字母，参考第一部分 b。		
针对中国数据，主要是 CN		

DOCUMENTNUMBER	
此项取值为标准化后的公开号，标准化格式只含数字，数字不超过 13 位	

DATE_PRODUCED	
适用于日期的规则，参考第一部分 c。	

DATE_PUB	
适用于日期的规则，参考第一部分 c。	

STATUS			
原始XML中的值	数据库中对应的值(查询条 件)	备注	
C	1	新建,中国数据都是 C	
D	2	删除	
R	3	替换	

T_BIB_PUB_REF

T_BIB_APP_REF

T_BIB_PRIORITY_DETAILS

PRIORITY_ATT_DATAFORMAT, PRIORITY_ATT_KIND,

PRIORITY_ATT_DATASOURCE, KIND, DATE_APP 与之前各个表的验证方式类

似。

T_BIB_INVENTION_TITLE

ATT_LANG, ATT_PROCESSING_TYPE, ATT_CREATOR, ATT_CREATE_DATE,

ATT_DATA_SOURCE 字段请参考前面的验证方法。

TITLE 字段是 XML 片段不适合使用数据库质检。

T_BIB_ASSIGNEE

ATT_LANG, ATT_DATAFORMAT, ATT_SOURCEDB,

ATT_PROCESSING_TYPE, ATT_CREATOR 参考前面的验证方法。

ADDRESSBOOK 字段是 XML 片段不适合用数据库质检。

CODE_ASSIGNEE	000000000000
---------------	--------------

T_BIB_APPLICANT

ATT_APPTYPE, ATT_DESIGNATION, ATT_DATAFORMAT,

ATT_DATASOURCE, ATT_STATUS, ATT_LANG, ATT_CREATE_DATE,

ATT_CREATOR, ATT_PROCESSING_TYPE, NATIONALITY, RESIDENCE 参考前

面的验证方法。

CODE_ORGANIZATION	10 位 0
-------------------	--------

CODE_CREATEDATE	8 位 0
-----------------	-------

CODE_CREATOR	2 位 0
在导出时产生	

T_BIB_INVENTOR

ATT_DATAFORMAT, ATT_DATASOURCE, ATT_CREATOR,

ATT_CREATE_DATE, ATT_PROCESSING_TYPE 参考前面的验证方法。

T_BIB_AGENT

ATT_LANG, ATT_PROCESSING_TYPE, ATT_CREATOR 参考前面的验证方

法。

T_BIB_MISC

DATE_NATIONAL_STAGE_ENTRY, DATE_PCT_PUBLICATION, DATE_PCT_APPLICATION 参考前面的验证方法。

第二步：标准化文本数据成品质检
(一) 机器全量质检：

A：将全部文本数据进行 DTD 标准化格式校验，符合标准化格式规范。
B：将全部文本数据进行 52 位文件名校验，符合文件名规范。

C：对全部文本数据的所有著录项目、摘要、摘要附图、权利要求书、说明书、说明书附图等规范化校验。

质检规则如下表所列：

PatentDocumentAndRelated		
lang	2 位小写国别字母 zh	country
docNumber	此项取值为标准化后的公开号，标准化格式只含数字，数字不超过 13 位	
kind	取公开类型值	
dateProduced	8 位日期型数字	
datePublication	8 位日期型数字	
status	中国数据该项值应全为 C	
BibliographicData		
lang	2 位小写国别字母 zh	country
PublicationReference		
dataFormat	dataFormat='standard', dataFormat='original' 必须出现；	
sourceDB	dataFormat='original' 时 sourceDB='national office' 必须出现；	
sequence	必选；按多值出现顺序排列 sequence=1, sequence=2, sequence=3...	

PublicationReference\base: DocumentID			lang	2 位小写国别字母 zh	WIPOST3Code	两位大写国别字母 CN	DocNumber	标准化格式只含数字，不超过 13 位	Kind	值为 1 位字母或 1 位字母+1 位数字	PublicAvailabilityDate\GazetteReference			date		8 位日期型数字	ApplicationReference			applType		必选：值为 10（发明）或 20（实用新型）
											</											

		creator	必须出现
		lang	必须出现；两位小写对应字母
InventionTitle			
	ClassificationDataSource	取值：人工分类,用大写字母H表示	
	ClassificationStatus	文献中首次给出的数据是初始数据,即使某一专利局以基本版	
	GeneratingOffice	2 位大写字母	
	base: ActionDate	8 位数字	
	ClassificationValue	取值：发明用大写字母I表示;非发明用大写字母N表示	
	SymbolPosition	取值为大写字母F(第一位),大写字母L(在后位置)	
	Subgroup	只含数字;斜线后1-3位的阿拉伯数字表示,取值为1-999,如	
	MainGroup	只含数字;与前面的小类号和后默认"00",组成完整	
	Subclass	1 位大写字母;大写字母A-Z	
	MainClass	2 位阿拉伯数字,01-99;	
	Section	1 位大写字母;用大写字母A-H表示;	
	ClassificationLevel	字母C(基本版)、A(高级版)和S(小类)	
	IPCVersionDate	8 位数字,20060101	
ClassificationIPCRDetails			
	FurtherClassification	dataFormat="original"时,必选	
	MainClassification	dataFormat="original"时,必选	
ClassificationIPC			
	Date	8 位日期型数字	
	DocNumber	标准化格式只含数字	
	WIPOST3Code	2 位大写字母	
	sourceDB	dataFormat='original'时,sourceDB='national'	

Priority		
	Date	8 位数字
	WIPOST3Code	2 位大写字母
	lang	2 位小写字母
ParentDocument		
Division		
	Date	8 位数字
GazetteNumber		
	lang	2 位小写字母
	country	2 位大写字母
GazetteReference		
	country	2 位大写字母
RelatedPublicationDoc	lang	2 位小写字母
BibliographicData		
	sourceDB	必须出现;
	dataFormat	必须出现; 取值 standard、original、other 等
	creator	必须出现;
	lang	必须出现; 两位小写对应字母
Abstract		
	sourceDB	必须出现;
	dataFormat	必须出现; 取值 standard、original、other 等
	creator	必须出现
	lang	必须出现; 两位小写对应字母
AssigneeDetails		
	sourceDB	必须出现; 数据来源取值如 EPO, national office,
	dataFormat	必须出现; 取值 standard、original、other 等

CorrespondenceAddress		
	sourceDB	必须出现;
	dataFormat	必须出现; 取值 standard、original、other 等
	lang	必须出现; 两位小写对应字母
	creator	必须出现;
	publicationMark	取值 0 (公布姓名 yes) 或 1 (不公布姓名 no)
InventorDetails		
	Residence	2 位大写国别字母
	Nationality	2 位大写国别字母
	creator	2 位 0
	createDate	8 位 0
	OrganizationCode	10 位 0
	sourceDB	必须出现;
	dataFormat	必须出现; 取值 standard、original、other 等
	lang	必须出现; 两位小写对应字母
	creator	必须出现;
Applicant		
Parties		
	FilingOffice	2 位大写字母
	base:Date	8 位日期型数字
	base:DocNumber	标准化格式只含数字
	base:WIPOST3Code	2 位大写字母
	sourceDB	dataFormat='original' 时 sourceDB='national office' 必须出现
	dataFormat	dataFormat='standard' 必须出现;
kind		national: 优先权国家与公开国家相同; international: 优先权国家与公开国家不同;

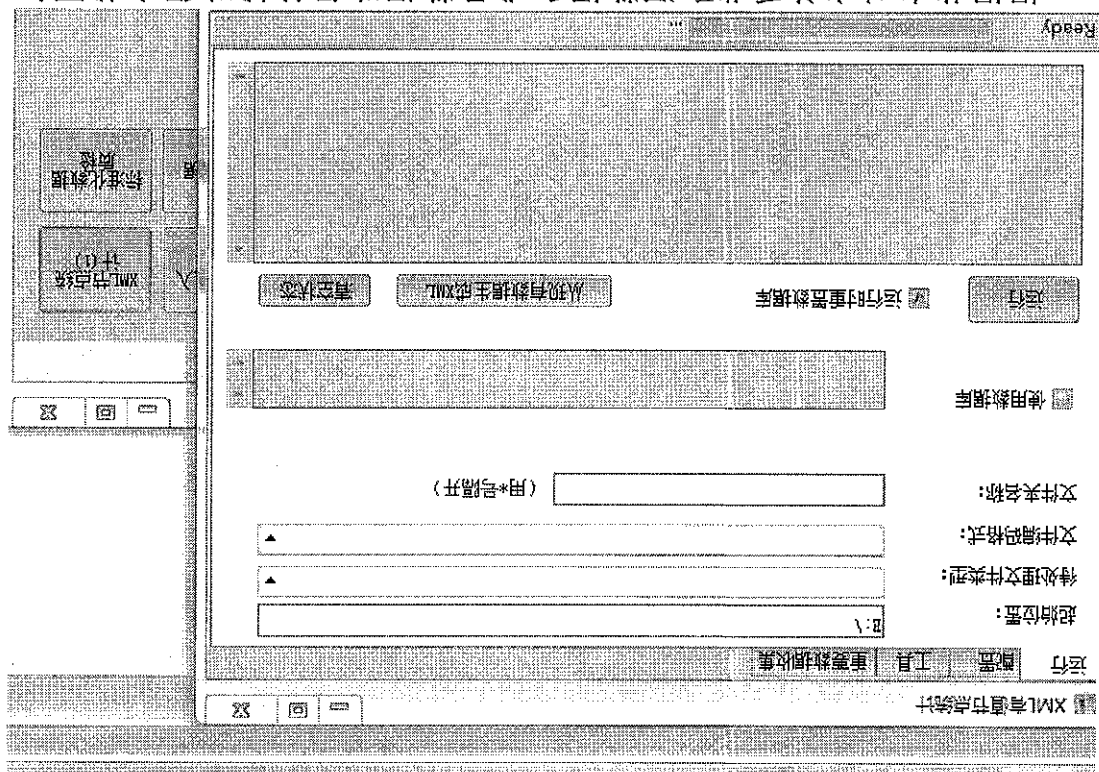
(二) 人工抽检:

		GazetteReference\Gaze teNumber@Date	8 位数字
		GazetteReference\lang	2 位小写字母
		GazetteReference\count ry	2 位大写字母
		Date	
PCTORRegionalPublicationData			
		NationalStageEntryDat e	8 位日期型数字
		Date	8 位日期型数字
PCTORRegionalApplicationData			
		base: Date	8 位日期型数字
NationalStageEntryDate			
		sourceDB	必须出现;
		dataFormat	必须出现; 取值 standard、original、 other 等
		lang	必须出现; 两位小写对应字母
		creator	必须出现;
AgentDetails			
		CustomerNumber	5 位数字或 5 位 0
		CustomerNumber	5 位 0
		sourceDB	必须出现;
		dataFormat	必须出现; 取值 standard、original、 other 等
		lang	必须出现; 两位小写对应字母
		creator	必须出现;

按本批次成品数据总量，对成品数据抽取比例为 12% 的样本产品，依据标准化格式转化规则进行著录项目、摘要、摘要附图、权利要求书、说明书、说明书附图等逐项质检。

(1) 全节点样例质检，确保生成的所有节点转换正确，程序辅助生成全节点样例数据。

A: 生成全节点样例数据的程序截图如下：



B: 根据节点对应关系获取源数据和成品数据路径的辅助程序截图如下：

下：

(一) 机器全量质检:

第三步：文本数据的索引文件检查：

[illegible]

成品数据文件截图如下:

20170915-1-001.ZIP
20170915-2-001.ZIP
20170915-3-001.ZIP
20170915-4-001.ZIP
20170915-5-001.ZIP
20170915-6-001.ZIP
20170915-7-001.ZIP
20170915-8-001.ZIP
20170915-9-001.ZIP

(1) 数据准备:

- A: 路径索引文件信息入库(路径索引文件所在文件夹全路径, 路径索引文件名, 索引文件行内容, Taskid(路径索引文件所在文件夹中国别层级上一层的文件名))
- B: 号单索引文件信息入库(号单索引文件所在文件夹全路径, 号单索引文件名, 每件专利标准公开四信息, Taskid(号单索引文件所在文件夹中国别层级上一层的文件名)), 成品文件 Status 状态, DateProduce 值, File 属性值) 关于 Taskid 需要在入库完成后将不是 Taskid 的值清除

(2) 索引文件与成品文件之间的质检项目:

(2-1) 质检已入库的三种索引文件信息是否重复的校验

路径索引的重复条件: 路径索引文件所在文件夹全路径 + 索引文件行内容

号单索引的重复条件: 号单索引文件所在文件夹全路径 + 每件专利标准公开四项信息

核验索引的重复条件: 核验索引文件所在文件夹全路径 + 每件专利标准公开四项信息

(2-2) 成品文件与索引一一对应的质检:

A、根据输入的截止到公开日文件夹路径(可多个)进行校验, 每个路径进行一次以下处理

B、根据截止到公开日文件夹路径, 加载数据库中三种索引的信息

C、遍历所有成品文件，获取三个信息（成品文件内容中标准公开四信息、成品文件全路径，成品文件的流水号上层文件夹全路径，与核验索引 File 属性值相同层级以下相对路径）

D、对于路径索引：使用 1-C 中的 1,3 项信息与数据库路径索引 1,3 项目信息进行匹配

对于号单索引：使用 1-C 中的 1,3 项信息与数据库号单索引 1,3 项目信息进行匹配

对于核验索引：使用 1-C 中的 1,3,7 项信息与数据库核验索引 1,3,4 项目信息进行匹配

匹配到的从内存中删除对应索引信息，未能匹配的属于对应索引中无成品文件信息。（生成三类异常日志名称：1、有成品无对应的路径索引；2、有成品无对应号单索引；3、有成品无对应核验索引）

E、完成 D 的校验后，内存中加载的索引信息仍然有内容的，属于有索引但没有成品文件。（生成三类异常日志名称：1、有路径索引无成品；2、有号单索引无成品；3、有核验索引无成品）

(3) 以核验索引为基准对数据库的质检项目：

A、在三索引文件与成品文件之间的质检正确的基础上（以截止到公开日文件夹路径为单位）可以进行与数据库之间的质检

B、根据截止到公开日文件夹路径，加载数据库中核验索引的信息

C、使用内存中核验索引的公开四信息到数据库中匹配。若匹配到

且 STATUS 一致的只有一件则正确，若不存在则为说明数据库不存在，若存在且 STATUS 有不一致则说明有一方的状态可能不正确，若存在且 STATUS 一致的有多件，需要追加 TASKID（值不是空白的，否则无法判断并记录日志）再次进行以上判断

注：本校验核索引表进行标记，对于正确的标记 1，不正确的标记

0。

生成三类异常日志名称：1、核验索引的公开四信息在数据库中不存
在；2、核验索引的公开四信息在数据库中 STATUS 有不一致；3、核验
索引的公开四信息在数据库中 STATUS 一致的有多件

索引文件校验程序的截图如下：

校验索引文件程序(V1.17.2.2001)

加工库连接信息: user id=di_data source=di_password=di_

索引文件类型: XML

Taskid: 1

成品文件关位置:

执行状态: 已经导入101条, 还剩余11条

开始

(二) 人工抽检

人工抽检：按本批次成品数据总量，对成品索引文件抽取比例为 12% 的样本产品，主要针对索引文件自身是否重复，索引文件与成品文件是否对应，索引文件与数据库内容是否对应。

质检数据示例截图如下：

路径索引、号单索引及核验索引文件截图：

- 20170915-INDEX-001.TXT
- 20170915-INDEX-A-001.XML
- 20170915-INDEX-A8-001.XML
- INDEX-CN-20170915-A-001.XML
- INDEX-CN-20170915-A8-001.XML

路径索引内容截图如下:

X

[illegible][illegible]

号单索引内容截图如下:

9.2 标准化图像数据成品检测

流程图如下：



核索引内容截图如下：

count	status	datePublished	path
1001	1001	1001	1001
1002	1002	1002	1002
1003	1003	1003	1003
1004	1004	1004	1004
1005	1005	1005	1005
1006	1006	1006	1006
1007	1007	1007	1007
1008	1008	1008	1008
1009	1009	1009	1009
1010	1010	1010	1010
1011	1011	1011	1011
1012	1012	1012	1012
1013	1013	1013	1013
1014	1014	1014	1014
1015	1015	1015	1015
1016	1016	1016	1016
1017	1017	1017	1017
1018	1018	1018	1018
1019	1019	1019	1019
1020	1020	1020	1020
1021	1021	1021	1021
1022	1022	1022	1022
1023	1023	1023	1023
1024	1024	1024	1024
1025	1025	1025	1025
1026	1026	1026	1026
1027	1027	1027	1027
1028	1028	1028	1028
1029	1029	1029	1029
1030	1030	1030	1030
1031	1031	1031	1031
1032	1032	1032	1032
1033	1033	1033	1033
1034	1034	1034	1034
1035	1035	1035	1035
1036	1036	1036	1036
1037	1037	1037	1037
1038	1038	1038	1038
1039	1039	1039	1039
1040	1040	1040	1040
1041	1041	1041	1041
1042	1042	1042	1042
1043	1043	1043	1043
1044	1044	1044	1044
1045	1045	1045	1045
1046	1046	1046	1046
1047	1047	1047	1047
1048	1048	1048	1048
1049	1049	1049	1049
1050	1050	1050	1050
1051	1051	1051	1051
1052	1052	1052	1052
1053	1053	1053	1053
1054	1054	1054	1054
1055	1055	1055	1055
1056	1056	1056	1056
1057	1057	1057	1057
1058	1058	1058	1058
1059	1059	1059	1059
1060	1060	1060	1060
1061	1061	1061	1061
1062	1062	1062	1062
1063	1063	1063	1063
1064	1064	1064	1064
1065	1065	1065	1065
1066	1066	1066	1066
1067	1067	1067	1067
1068	1068	1068	1068
1069	1069	1069	1069
1070	1070	1070	1070
1071	1071	1071	1071
1072	1072	1072	1072
1073	1073	1073	1073
1074	1074	1074	1074
1075	1075	1075	1075
1076	1076	1076	1076
1077	1077	1077	1077
1078	1078	1078	1078
1079	1079	1079	1079
1080	1080	1080	1080
1081	1081	1081	1081
1082	1082	1082	1082
1083	1083	1083	1083
1084	1084	1084	1084
1085	1085	1085	1085
1086	1086	1086	1086
1087	1087	1087	1087
1088	1088	1088	1088
1089	1089	1089	1089
1090	1090	1090	1090
1091	1091	1091	1091
1092	1092	1092	1092
1093	1093	1093	1093
1094	1094	1094	1094
1095	1095	1095	1095
1096	1096	1096	1096
1097	1097	1097	1097
1098	1098	1098	1098
1099	1099	1099	1099
1100	1100	1100	1100

生物保藏: BID

按照条约第 19 条修改的权利要求书: AMDCLA

生物序列表: GENE

说明书附图: DRA

说明书: DES

权利要求书: CLA

扉页: BIB

PDF 页签对应关系如下:

质检规则:

文件能否正确打开。

C: 对全部 PDF 文件数据的所有标签信息进行程序读取解析, 及 PDF

B: 将全部 PDF 文件数据进行 52 位文件名校验, 符合文件名规范。

标准化格式规范。

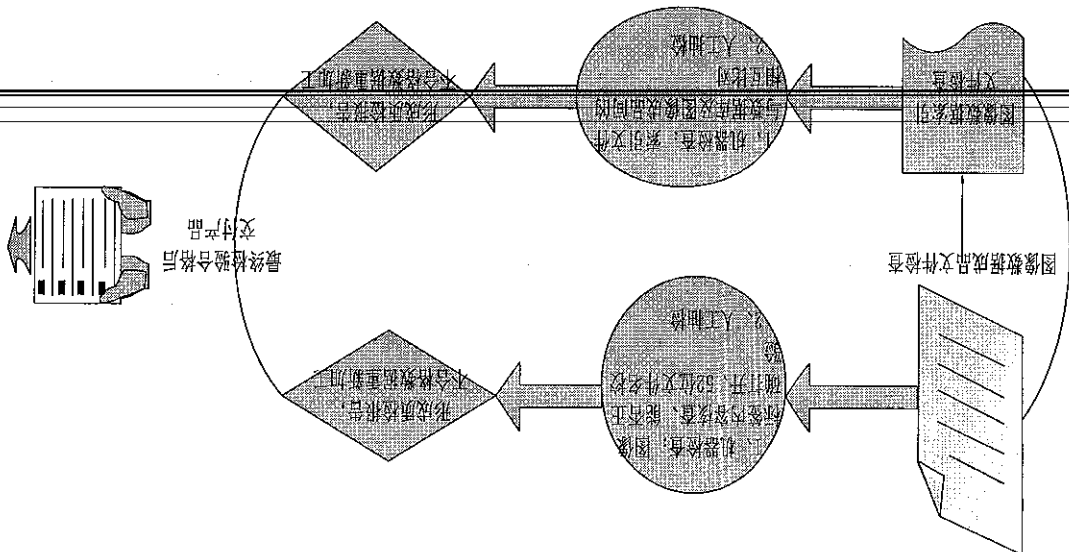
A: 将 PDF 文件生成相应的 XML 文件进行 DTD 标准化格式校验, 符合

(一) 机器全量质检:

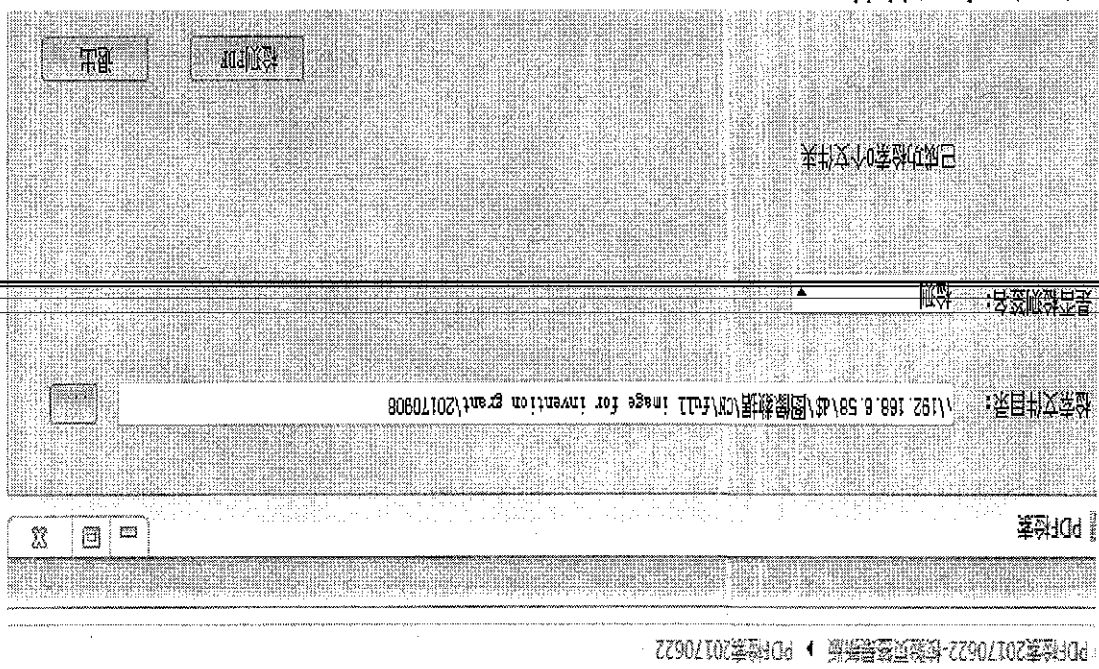
第一步: 图像数据的成品文件质检:

该成品检测分两步:

流程说明如下:



PDF 成品文件数据检测程序截图如下:



(二) 人工抽检:

按本批次成品数据总量,对成品数据抽取比例为15%的样本产品,查看样本PDF文件生成相应的XML文件内容是否正确;PDF文件数据52位文件名及页签信息是否正确。

PDF压缩包截图如下:



PDF文件生成相应的XML文件内容截图如下:

[illegible]

校验 PDF 索引文件与标准化文本文件信息是否对应的程序截图如下：

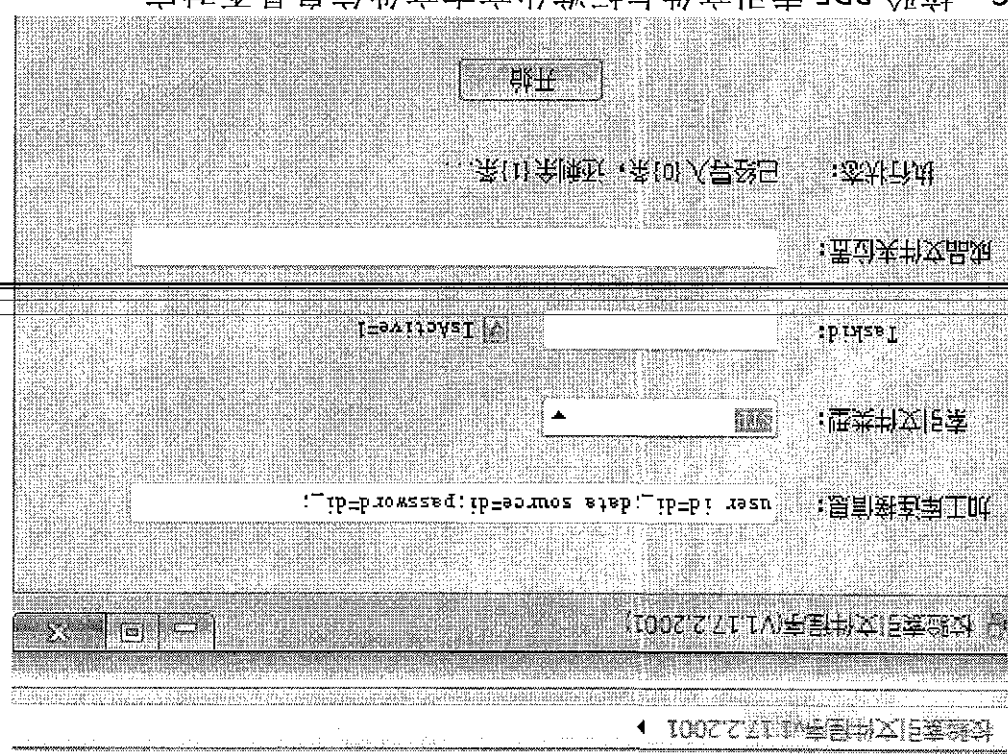
志。

与标准化文本数据中的对应号日信息进行对比，将差异性给出相应日

```
ApplicationDate="20060926" PublicationNum="CN101501640B"
PublicationDate="20170908" />
<Doclist ApplicationNum="CN200680039792.7"
```

读取索引文件中的号日信息：

C：校验 PDF 索引文件与标准化文本文件信息是否对应



按本批次成品数据总量，对成品数据抽取比例为 15% 的样本产品，
主要针对 PDF 索引文件内容是否正确，PDF 索引文件与 PDF 数据文件是否
对应。

PDF 路径索引文件质检点：52 位文件夹名、52 位文件名。

示例：

\\CN11200600039792CN00001015016400BPIDZH20170908CN00\CN1
12006000039792CN00001015016400BPDFZH20170908CN00#.PDF
\\CN11200600039792CN00001015016400BPIDZH20170908CN00\CN1
12006000039792CN00001015016400BPIDZH20170908CN00.XML

PDF 号单索引文件质检点：申请号、申请日、公开号、公开日。

示例：

<Doclist ApplicationNum="CN200680039792.7"
ApplicationDate="20060926" PublicationNum="CN101501640B"
PublicationDate="20170908"/>

PDF 路径索引及 PDF 号单索引截图：



20170908-INDEX-B-001.XML

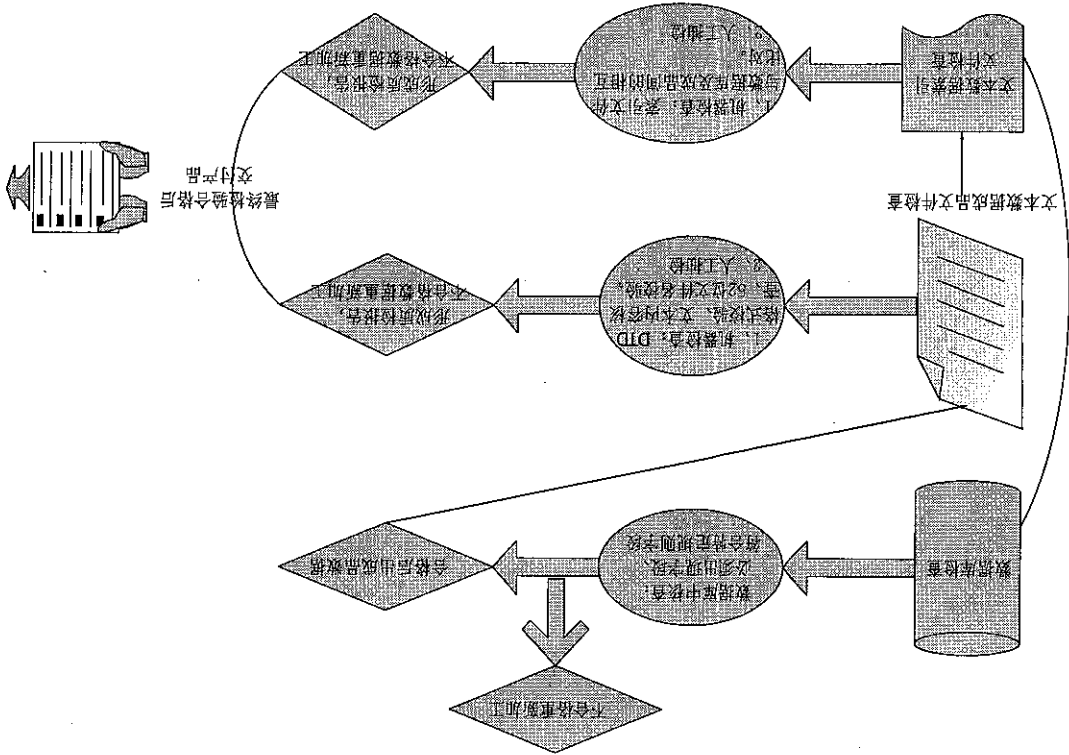
20170908-INDEX-001.TXT

(二) 人工抽检：

数据索引文件号对比程序	
数据路径：	\\192.168.0.142\全文图像目录及索引结构
号单路径：	\\192.168.0.71\加工数据\号单比对\AN_PN_INDEX_20170627
年份：	
数据类型：	RMSQ FULLIMAGE_INVENTION_GRANT
15h015	
开始	退出

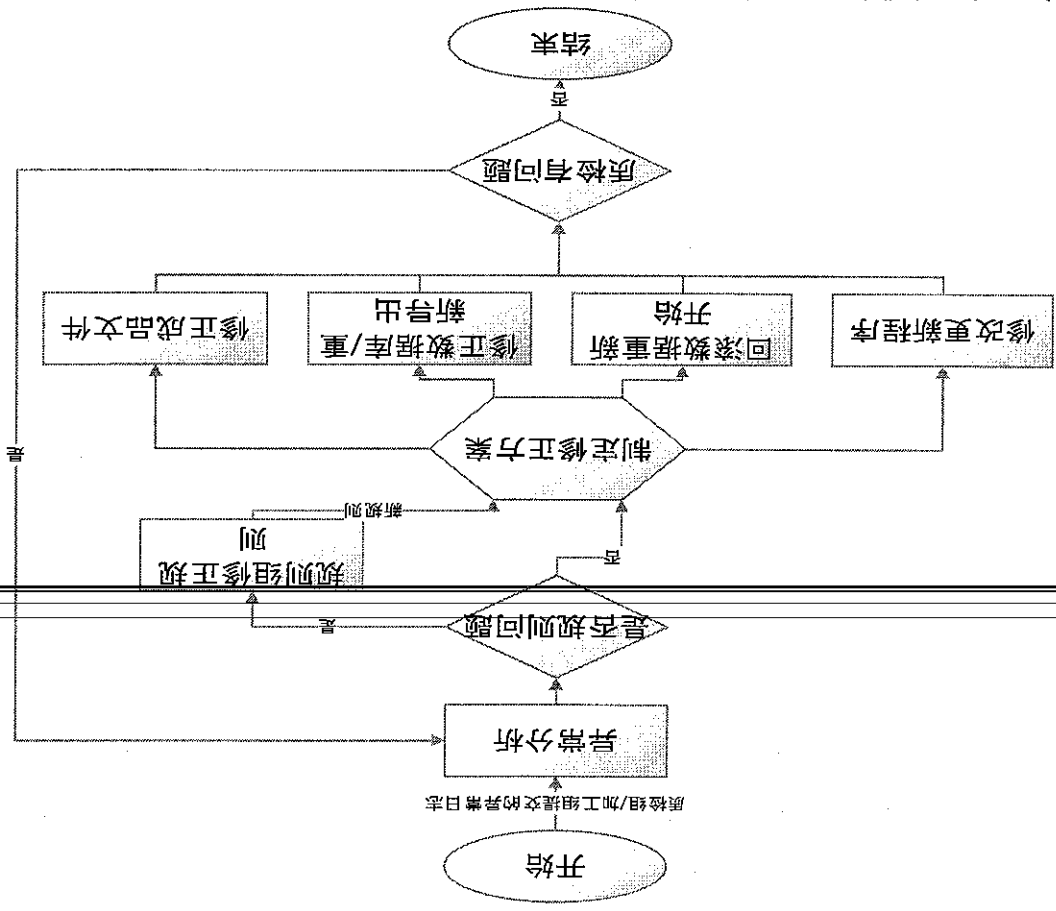
10 质检问题处理

1、针对文本数据标准化加工中的数据成品进行内部质量检测流程：



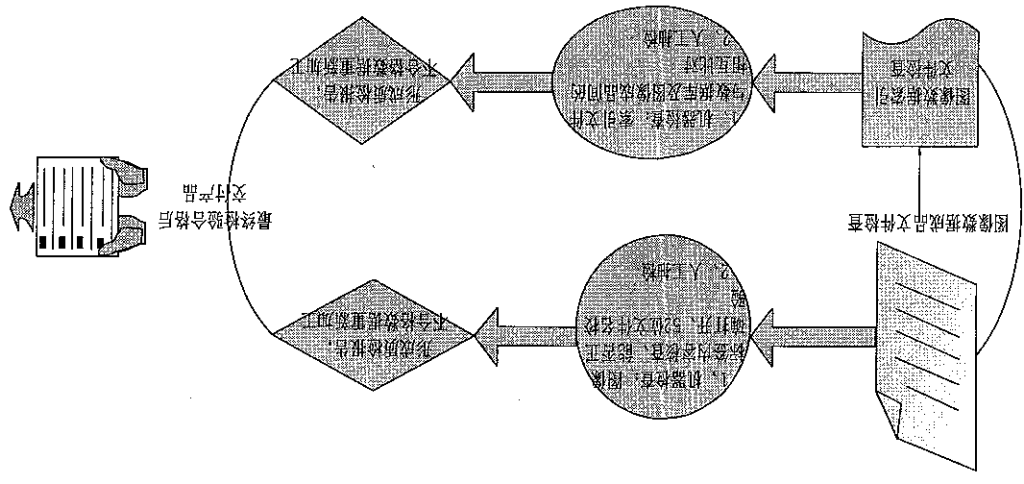
对于质检不合格的各类当前批次文本数据，都将返回到重新加工步骤，同时形成相应的质检问题报告，在错误批次数据重新加工之前，会先

根据此质检问题报告对导致问题的加工规则或加工程序进行修正和优化，即进入到问题分析与处理步骤，具体流程如下：

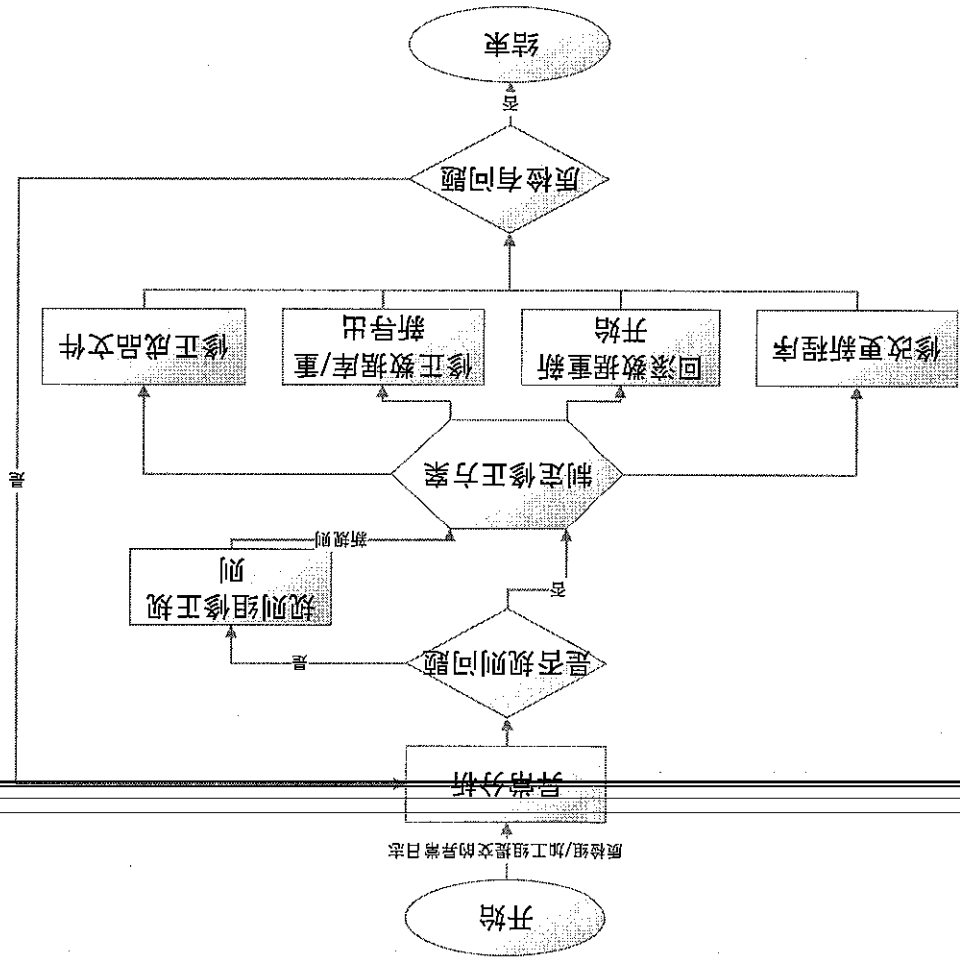


在对加工规则或加工程序进行修正后，此批次文本数据开始进入重新加工及其后续加工步骤，直到通过后续所有内部质检流程后，达到内部质量检测标准，本批次数据的质检工作完成。

2、针对图像数据标准化加工中的数据成品进行内部质量检测流程：



对于质检不合格的各类当前批次图像数据，都将返回到重新加工步骤，同时形成相应的质检问题报告，在错误批次数据重新加工之前，会根据此质检问题报告对导致问题的加工规则或加工程序进行修正和优化，即进入到问题分析与处理步骤，具体流程如下：



在对加工规则或加工程序进行修正后，此批次图像数据开始进入重新加工及其后续加工步骤，直到通过后续所有内部质检流程后，达到内部质量检测标准，本批次数据的质检工作完成。

9
5

9
5