

中国专利全文文本数据

（一）数据名称

中国专利全文文本数据

（二）文献语种

中文

（三）更新周期

周更新

（四）数据内容及格式

1、基本内容

申请信息、公开信息、权利要求书信息、说明书信息、说明书附图信息

2、特色内容

权利要求书：独立权利要求项、从属权利要求项标引，权利要求类型

说明书：说明书五部分标引

其他：权利要求书关键词，说明书关键词。

3、数据格式

XML、GIF

（五）文件结构

1、概述

本数据的文件存储层级结构包括三层文件夹，实体数据存储在第三层文件夹下。

第一层文件夹：数据类别。数据类别由数据内容描述（小写英文字母）+语种代码（两位大写英文字母）表示。例如：full text(ZH)。

第二层文件夹：文献种类描述，由小写英文字母组成，例如：invention publication。

第三层文件夹：公开日期，8位数字，YYYYMMDD（4位年+2位月+2位日），例如：20160330。

2、相关说明

1) 数据内容描述说明

bibli and abstract: 著录项目与文摘数据

full text: 全文文本

full image: 全文图像

legal status: 法律状态

enriched citation: 引文

family: 同族

reexamination: 复审

invalidation: 无效

verdict: 判决书

2) 语种代码说明

ZH: 表示中文数据

EN: 表示英文数据

3) 文献种类描述说明:

invention publication: 发明专利申请

invention grant: 发明专利授权

utility model: 实用新型专利授权

design: 外观设计专利授权

(六) 数据说明

1、文件包层级结构说明

示例: YYYYMMDD-N1.ZIP/C1NNNNNNNNNK1/ C1NNNNNNNNNK1. XML

YYYYMMDD-N1.ZIP

YYYYMMDD 为公开(公告)日, 如 20160330; N1 表示任一公开(公告)日的
数据按容量是否超过 2G 分隔的压缩包个数, 每个压缩包的数据容量以 2G 进行划
分, 超过 2G 后顺序排列, 如 20160330-002.ZIP 为 20160330 期的第二个 2G 压缩
包...以此类推。

C1NNNNNNNNNK1

C1为公开国别代码(二位大写字母), NNNNNNNNN为公开流水号, K1为文献

类型,C1NNNNNNNNK1为文件夹名称。该文件夹中包括:文件名称为C1NNNNNNNNK1.XML的全文代码化文件以及GIF格式的说明书附图。

2、主要数据元素说明

元素名称	元素内容
PublicationInfo	公开信息
ApplicationInfo	申请信息
Claims	权利要求书信息（包括独权项、从权项标引，权利要求类型）
Description	说明书信息（包括五部分标引）
Drawings	说明书附图信息
Keywords	权利要求书、说明书关键词