

中国发明专利授权公告标准化全文文本数据

(一) 数据编号

CN-TXTS-10-B

(二) 数据名称

中国发明专利授权公告标准化全文文本数据

(三) 文献语种

中文

(四) 更新周期

每周三更新

(五) 数据内容及格式

数据内容： 著录项目、摘要、权利要求书、说明书、说明书附图、索引信息。

数据格式： XML、TIFF

(六) 文件结构

1.概述

本数据的文件结构包括两层文件夹，实体数据存储在第一层文件夹下。

第一层文件夹：数据编号与数据名称。数据编号由国别代码(两位大写字母)+文献内容代码(三位大写字母)+文献内容属性代码(一位大写字母)+专利类型代码+文献种类代码组成；数据名称为数据中文全称。例如：CN-TXTS-10-B_中国发明专利授权公告标准化全文文本数据。

第二层文件夹：公布日期，8位数字，YYYYMMDD(4位年+2位月+2位日)，例如20160907。

2.相关代码说明

文献内容代码说明：

BIB： 著录项目数据

ABS： 摘要数据

TXT： 全文文本

IMG: 全文图像

文献内容属性代码说明:

O: 表示原始数据

T: 表示翻译数据

S: 表示标准化数据

专利类型代码说明:

10: 发明专利

20: 实用新型

30: 外观设计

文献种类代码说明:

A: 发明专利申请

B: 发明专利授权

U: 实用新型专利授权（韩国为实用新型专利申请）

Y: 实用新型专利授权（日本 2000 年以前，中国 1993-2010 年 4 月及韩国为实用新型专利授权）

S: 外观设计专利授权

(七) 数据说明

文件层级结构说明

第一层：存放三部分内容：一是“DATA”命名的文件夹，提供标准化全文文本数据文件的 ZIP 压缩包，压缩包命名方式为“YYYYMMDD-N1-N2.ZIP”，其中 YYYYMMDD 为公开（公告）日，如 20160907；N1 为文件夹流水号，每个流水号中不超过 2000 件专利文献，如本期数据超过 2000 件则按顺序流水编号，如 20160907-1-001.ZIP 为第一个 2000 件，20160907-2-001.ZIP 为第二个 2000 件... 以此类推；N2 表示每 2000 件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 20160907-1-002.ZIP 为 20160907 期的第一个 2000 件中的第二个 2G 压缩包... 以此类推。二是“INDEX”命名的文件夹，提供数据配套的索引文件，包括单索引文件和内容比对索引文件；三是“公开（公告）日-INDEX_流水号”命名的文件，提供数据配套的路径

Parties	当事各方
ApplicantDetails	申请人信息
InventorDetails	发明人信息
AgentDetails	代理人信息
AssigneeDetails	权利人信息
ExaminerDetails	审查员信息
Abstract	摘要
Description	说明书
Claims	权利要求
sequence-list-doc	序列表文件
date-pct-article-22-39-fulfilled	进入国家阶段日期（仅针对 PCT 申请）
pct-or-regional-filing-data	PCT 国际申请有关信息
pct-or-regional-publishing-data	PCT 国际申请国际公布有关数据