

美国专利申请公布全文文本数据

(一) 数据编号

US-PA-TXT0-10-A

(二) 数据名称

美国专利申请公布全文文本数据

(三) 文献语种

英文

(四) 更新周期

周更新

(五) 数据内容及格式

数据内容：包括著录项目、摘要、权利要求书、说明书、说明书附图、复杂单元（表格、数学式、化学结构和基因序列数据）。文档为XML格式，描述xml的DTD依照授权专利ICE4.3版本。表和序列数据使用CALS标记，数学表达式使用包括MathML和外部引用文件Mathematica Notebook (NB) 表示。化学式使用CDX和MOL文件表示，图像，数学式和化学式也以外部引用CCITT6.0标准的TIFF格式表示。每周补充超长序列（超过300页）或超长表格（超过200页），是ZIP文件。

数据格式：

数据种类	数据格式
文本数据	XML
图形数据	TIF
表和序列	CALS 标记
数学表达式	MathML、Mathematica Notebook (NB)、TIF
化学式	CDX、MOL、TIF

(六) 文件结构

1. 概述

本数据的文件结构包括一层文件夹，实体数据存储的第二层文件夹内。

第一层文件夹：数据编号与数据名称，例如：US-PA-TXTO-10-A_美国专利申请公布全文文本数据。

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 20141120。

2. 数据编号说明

编号方式：

国别代码-业务名称代码-数据资源名称代码-数据内容属性-专利类型代码-专利文献种类代码；

国别代码-业务名称代码-数据资源名称代码-数据内容属性-数据种类代码；

国别代码-业务名称代码-数据资源名称代码-数据内容属性-集成电路设计数据种类代码；

代码说明：

代码内容	编码方式	编码说明	参考依据
国别	两位大写字母	CN：中国 EP：欧洲专利局 JP：日本 KR：韩国 US：美国	WIPO ST.3 标准
业务名称	两位大写字母	PA：专利 TM：商标 GI：地理标志 IC：集成电路布图设计	《知识产权基础信息数据规范（试行2020版）》
数据资源名称	2-4位大写字母	BIB：著录项目 ABS：摘要 TXT：全文文本 IMG：全文图像 PUB：公告 PRS：法律状态 FIFT：日本 FI、Fterm 分类 USPC：美国专利分类 ECLA：欧洲 ECLA 分类 ICR：国际专利分类 DEC：决定	《知识产权基础信息数据规范（试行2020版）》

		RE: 复审案件 IN: 无效宣告案件 CIT: 引文	
数据内容属性	一位大写字母	0: 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	两位数字	10: 发明专利 20: 实用新型 30: 外观设计	
专利文献种类	一位大写字母	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;	
数据种类	2-4位大写字母	FILE: 表示为数据; DIC: 表示为字典、定义、说明文件等;	
集成电路布图设计数据种类	2-4位大写字母	ER: 专有权公告; ERA: 专有权事务公告; TER: 专有权终止公告; RE-RE: 复审及撤销案件;	

(七) 数据说明

主要数据元素说明:

元素名称	元素内容
us-patent-application	专利申请信息(包括语言、文档号、状态、id码、国家等)
us-bibliographic-data-application	专利申请著录项目数据
publication-reference	公开信息(包括国别、公开号、文献类型代码、公开日期)
application-reference	申请信息(包括申请类型、国别、申请号、申请日期)

us-application-series-code	美国专利申请序列代码
classifications-ipcr	国际专利分类数据 8 版
classification-national	美国国家分类
invention-title	发明标题
us-related-documents	美国在先及其他文档信息
parties	当事人信息（申请人信息）
assignees	专利权人信息
abstract	摘要
drawings	附图信息
description	说明书
id	Id 标签
description-of-drawings	附图说明
us-claim-statement	美国权利要求声明
claims	权利要求信息