

# 分词器操作手册

## 1. 分词器部署

按照《安装部署文件》对分词器进行部署。

## 2. 添加文章

左上角选择“添加文章”，选择当前已有类型或新增类型，并上传 excel 文件，文件第一行为表头，将需要标注的字段列命名为“正文”即可：

The screenshot displays the '政务分词器' (Government Word Segmentation) application interface. On the left, a sidebar contains a '添加文章' (Add Article) button and a list of articles. The main content area shows a table with columns for '标题' (Title), '序号' (Serial Number), and '发刊时间' (Publication Time). The right-hand panel, titled '标注新关键词' (Tag New Keywords), lists various keywords for tagging, such as '道路运输及站场管理规定', '交通运输部', and '中华人民共和国交通运输部'. A '保存' (Save) button is visible at the bottom of the right panel.

标题	序号	发刊时间
国务院公报中华人民共和国交通运输部令（2023年第12号）交通运输部关于修改《道路货物...	国务院公报-180	2024-1-30
文号	2024年第3号（总号：1830）	

中华人民共和国交通运输部令2023年第12号《交通运输部关于修改〈道路货物运输及站场管理规定〉的决定》已于2023年11月1日经第24次部务会议通过，现予公布，自公布之日起施行。部长 李小鹏  
2023年11月10日 交通运输部关于修改《道路货物运输及站场管理规定》的决定交通运输部决定对《道路货物运输及站场管理规定》（交通运输部令2022年第30号）作如下修改：一、将第六十一条修改为：“违反本规定，有下列行为之一的，由交通运输主管部门责令停止经营；违法所得超过1万元的，没收违法所得，处违法所得1倍以上5倍以下的罚款；没有违法所得或者违法所得不足1万元的，处3000元以上1万元以下的罚款，情节严重的，处1万元以上5万元以下的罚款；构成犯罪的，依法追究刑事责任：（一）未按规定取得道路货物运输经营许可证，擅自从事道路普通货物运输经营的；（二）使用失效、伪造、变造、被注销等无效的道路运输经营许可证从事道路普通货物运输经营的；（三）超越许可的事项，从事道路普通货物运输经营的。”二、删去第六十三条第三款。三、将第六十六条中的“处1万元以上3万元以下的罚款”修改为“处3000元以上3万元以下的罚款”。本决定自公布之日起施行。《道路货物运输及站场管理规定》根据本决定作相应修正，重新公布。道路货物运输及站场管理规定（2005年6月16日交通部发布 根据2008年7月23日《交通运输部关于修改〈道路货物运输及站场管理规定〉的决定》第一次修正 根据2009年4月20日《交通运输部关于修改〈道路货物运输及站场管理规定〉的决定》第二次修正 根据2012年3月14日《交通运输部关于修改〈道路货物运输及站场管理规定〉的决定》第三次修正 根据2016年4月11日《交通运输部关于修改〈道路货物运输及站场管理规定〉的决定》第四



### 3. 数据标注

#### 3.1. 数据类型选择

添加完毕文件后，在菜单下选择需要标注的数据列表：



#### 3.2. 序号范围选择

按照实际情况选择需要标注的数据序号范围：



### 3.3. 显示标注结果

选中需要标注的文章，点击“显示标注结果”，可看到当前文章中已被标注的关键词，已标注的关键词不必进行重复标注：



### 3.4. 标注

在文章的内容中进行标注，鼠标左键选中正文中的目标词，点击右键“添加新关键词”或使用快捷键 Ctrl+k/Shift+s/enter 对选中词进行标注：



右侧“标注新关键词”栏中，显示紫色的关键词并未被存储，建议标注 20 个词左右时点击“保存”按钮进行存储，防止出现进行大量标注之后存储失败导致需要重新标注的情况（若显示存储失败，再

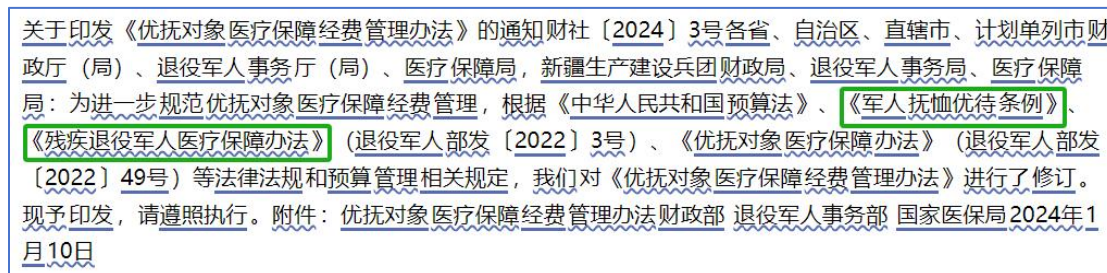


### 3.6. IK 分词验证

文章标注的关键词存储后，大概 2~3 分钟，关键词会被添加至 IK 分词中，点击“IK 分词验证”，则可看到当前的分词情况：



横线和波浪线的作用是为了分词，也就是说对 IK 分词器来说，连续横线上的为一个分词，连续波浪线上的为一个分词；如下图中的“军人抚恤优待条例”没有被标注，就被分词为“军人”“抚恤优待”“条例”三个词；如“残疾退役军人医疗保障法”已被标注，则只将其划分为一个词。可以在“IK 分词验证”效果下，查看有哪些分词错误并添加新分词对效果进行更正（IK 分词验证会在新分词保存 1-3 分钟后生效）：



### 3.7. 导出分词

标注完成后，选择菜单中的导出分词，所有标注词将会以 txt 格式导出至本地。

## 4. 状态列表

### 4.1. 未标注

列表显示所有未进行手动标注的文章：

The screenshot shows the '政务分词器' (Government Word Segmentation) interface. On the left, there is a sidebar with a list of documents, including '国务院部门文件 (11035)'. The main area displays a table of unannotated articles. The table has columns for '标题' (Title), '序号' (Serial Number), '发文机关' (Issuing Agency), '发文字号' (Document Number), '来源' (Source), '主题分类' (Topic Classification), '公文种类' (Document Type), '成文日期' (Effective Date), and '发布日期' (Release Date). The first article is '国家药监局综合司关于印发药品监管人工智能典型应用场景清单的通知' (Notice of the General Administration of China Food and Drug Administration on Issuing a List of Typical Application Scenarios for Artificial Intelligence in Drug Supervision). Below the table, there is a detailed description of the notice. On the right, there is a panel for '标注新关键词' (Annotate New Keywords) and '关键词记录' (Keyword Record).

标题	序号	发文机关
国家药监局综合司关于印发药品监管人工智能典型应用场景清单的通知	国务院部门文件-4	国家药监局
发文字号	药监综函〔2024〕313号	来源 国家药监局网站
主题分类	市场监管、安全生产监管\食品药品...	公文种类 通知
成文日期	2024年06月13日	发布日期 2024-06-19

### 4.2. 标注中

若文章未完成标注且有标注词，则可在“标注中”列表查看未标注完成的文章，下次可再进入该文章进行标注：

