

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'. The interface also includes a search bar, a '重新分词' (Re-segment) button, and a '显示标注结果' (Show Tagging Results) button.

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，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件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

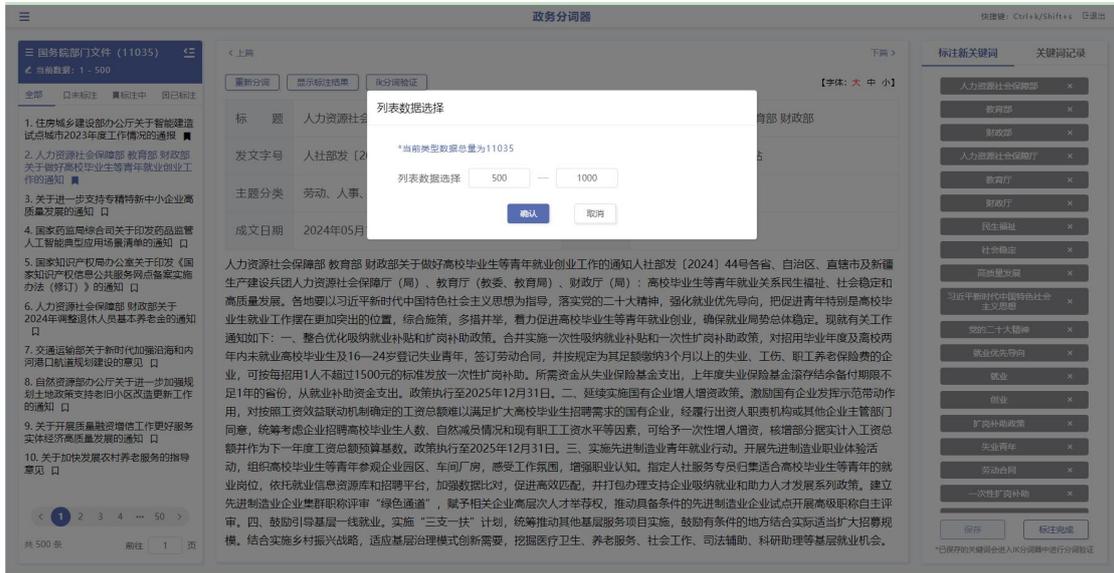
3.1. 数据类型选择

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

次点击存储按钮进行尝试)：



标注完成一篇文章时，点击“标注完成”，未存储的关键词会被存储，文章状态被标记为“已标注”。

3.5. 关键词记录

右侧“关键词记录”列表中，可查看关键词标注时间等情况：

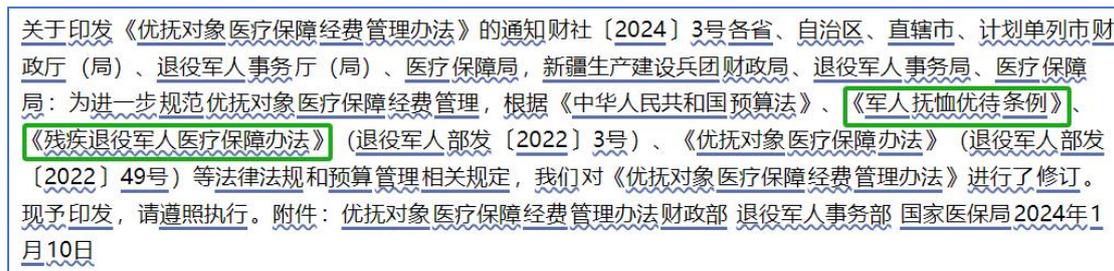


3.6. IK 分词验证

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 Tool) interface. On the left, there is a sidebar with a list of articles under the heading '国务院部门文件 (11035)'. The '未标注' (Unannotated) tab is selected. The main area displays a table of article details for the selected article: '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'.

标题	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		
序号	国务院部门文件-4	发文机关	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

Below the table, there is a text block containing the full content of the notice, starting with '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药监综函〔2024〕313号各省、自治区、直辖市和新疆生产建设兵团药品监督管理局，局机关各司局、各直属单位：为贯彻落实党中央、国务院关于开展“人工智能+"行动的决策部署，落实《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》《新一代人工智能发展规划》《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》《药品监管网络安全与信息化建设“十四五”规划》等工作要求，有效促进“人工智能+"行动在药品监管领域的实践探索，统筹推进人工智能场景创新，更好支撑高水平监管和高质量发展，现将《药品监管人工智能典型应用场景清单》（以下简称《清单》）印发给你们，请结合实际，积极探索实施。有关事项通知如下：一、工作目标《清单》列出了15个具有引领示范性的、有发展潜力的、针对工作痛点的、需求较为迫切的应用场景，旨在推动人工智能技术在药品监管领域的研究探索，以促进人工智能与药品监管深度融合为主线，规范和指导各级药品监管部门开展人工智能技术研究应用，引导资源聚焦，推动人工智能赋能药监系统；同时也为其他科研机构、技术公司和药品企业相关研究应用提供参考和指导。二、组织实施各单位应充分发挥主观能动性，结合人工

On the right side of the interface, there are buttons for '重新分词', '显示标注结果', and '分词验证'. At the bottom right, there are buttons for '保存' (Save) and '标注完成' (Annotation Complete).

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：

政务分离器

Ctrl+k/Shift+e 退出

国务院部门文件 (11035)

当前数据: 1 - 500

全部 未标注 标注中 已标注

1. 住房和城乡建设部办公厅关于智能建造试点城市2023年度工作情况的通报

2. 人力资源社会保障部 教育部 财政部 关于做好高校毕业生等青年就业创业工作的通知

标 题	住房和城乡建设部办公厅关于智能建造试点城市2023年度工作情况的通报	发文机关	住房和城乡建设部
发文字号	建办市函〔2024〕209号	来 源	住房和城乡建设部网站
主题分类	城乡建设、环境保护/城市规划	公文种类	通报
成文日期	2024年06月11日	发布日期	2024-06-20

住房和城乡建设部办公厅关于智能建造试点城市2023年度工作情况的通报建办市函〔2024〕209号各省、自治区住房和城乡建设厅，直辖市住房和城乡建设（管）委，新疆生产建设兵团住房和城乡建设局：为贯彻落实全国住房和城乡建设工作会议精神，按照住房和城乡建设部等部门《关于推动智能建造与建筑工业化协同发展的指导意见》（建市〔2020〕60号）等有关规定，我对24个智能建造试点城市（以下简称试点城市）2023年度工作情况开展了总结评估。现将有关情况通报如下。一、试点工作总体进展试点开展以来，各试点城市紧密围绕贯彻落实党的二十大精神

共 2 页 前往 1 页

标注新关键词 关键词记录

- 住房和城乡建设部
- 住房和城乡建设局
- 关于推动智能建造与建筑工业化协同发展的指导意见
- 建筑工业化
- 中央经济工作会议精神
- 建筑业转型
- 政策支持力度
- 高新技术企业
- 建筑信息模型
- BIM
- 建筑机器人

保存 标注完成

已保存的关键词进入关键词管理中进行分类验证

4.3. 已标注

“标注完成”的文章可在“已标注”列表中进行查看：

政务分离器

Ctrl+k/Shift+e 退出

国务院部门文件 (11035)

当前数据: 1 - 500

全部 未标注 标注中 已标注

1. 住房和城乡建设部办公厅关于智能建造试点城市2023年度工作情况的通报

2. 人力资源社会保障部 教育部 财政部 关于做好高校毕业生等青年就业创业工作的通知

标 题	住房和城乡建设部办公厅关于智能建造试点城市2023年度工作情况的通报	发文机关	住房和城乡建设部
发文字号	建办市函〔2024〕209号	来 源	住房和城乡建设部网站
主题分类	城乡建设、环境保护/城市规划	公文种类	通报
成文日期	2024年06月11日	发布日期	2024-06-20

住房和城乡建设部办公厅关于智能建造试点城市2023年度工作情况的通报建办市函〔2024〕209号各省、自治区住房和城乡建设厅，直辖市住房和城乡建设（管）委，新疆生产建设兵团住房和城乡建设局：为贯彻落实全国住房和城乡建设工作会议精神，按照住房和城乡建设部等部门《关于推动智能建造与建筑工业化协同发展的指导意见》（建市〔2020〕60号）等有关规定，我对24个智能建造试点城市（以下简称试点城市）2023年度工作情况开展了总结评估。现将有关情况通报如下。一、试点工作总体进展试点开展以来，各试点城市紧密围绕贯彻落实党的二十大精神

共 2 页 前往 1 页

标注新关键词 关键词记录

- 住房和城乡建设部
- 住房和城乡建设局
- 关于推动智能建造与建筑工业化协同发展的指导意见
- 建筑工业化
- 中央经济工作会议精神
- 建筑业转型
- 政策支持力度
- 高新技术企业
- 建筑信息模型
- BIM
- 建筑机器人
- 评估评价
- 骨干企业
- 示范项目
- 试验做法
- 创新赋能
- 专家评审
- 成果分析

保存 标注完成

已保存的关键词进入关键词管理中进行分类验证