

「吾与点」帮助中心

1. 「吾与点」平台简介

[「吾与点」智能标注平台](#)是北京大学数字人文研究中心研发的通用智能标注平台。该平台支持对自由文本的智能化结构整理、实体标注、关系标注和知识图谱自动生成功能，并导出 Brat、Doccano、Gephi、Echarts、BIOES 等多种常用数据格式。平台集成多种深度学习 AI 模型，并通过精巧的交互设计实现高效的人机协作。“吾与点”当前在文言文本上实现了实体识别、关系抽取和图谱自动生成，在现代汉语和英语文本上实现了实体识别。未来将支持处理更多语言。

2. 「吾与点」平台的目标和解决方案

目标：打造用户体验出色的智能数据生成平台，能够支持多用户使用和多种语言环境，完成智能数据生成的全系列任务，处理复杂的超长时段历时语料，并对接其他多种数据分析平台。

解决方案：使用 AI 训练模型对 33 亿字的古汉语数据进行训练，实现语言文本的自动标点、自动实体识别和关系提取。建设智能标注平台，实现项目管理、文件组织和数据标注的一体化功能集成，便于团队合作。

3. 「吾与点」平台快速使用流程

3.1. 上传文本数据

首先，用户在项目管理页面上传自己的文本数据。通过「新建项目（图 1）—>打开项目（图 2）—>新建文档/导入文档（图 3 和图 4）—>打开文档（图 5）」的路径上传文本数据。

假如只有单篇文档需要标注，可以点击「新建文档」按钮，在弹出的对话框中输入文档的名称，点击「确认」按钮后，将会看到我们创建好的空文档出现在项目的文档列表中。（图 3）。

如果有多个文本需要批量上传，可以点击「导入文档」按钮（图 4）。「吾与点」平台目前仅支持导入 txt 格式的文本文件，暂不支持导入 Word、WPS 文件。选择好文件后，点击「开始导入」按钮，上传完成后，页面将自动刷新，刚才导入的文档同样会出现在项目的文档列表中。



图 1



图 2



图 3



图 4



图 5

3.2. 结构标注

打开文件后, 首先进入的是「结构标注」页面。如果需要为文本自动加上标点, 可以点击「自动标点」按钮 (图 6), 句号、逗号等标点均已自动添加到文本中, 如有标点标错之处, 用户可以点击鼠标将光标定位到出错处, 自行编辑文本修改或删除标点。(图 7)



图 9

3.3. 实体标注

随后进入「实体标注」页面，在该页面，用户可以对文本进行自动实体标注和抽取。点击「实体标注」页面工具栏中的「自动标注实体」功能，就可以实现对实体的自动标注。（图 10）自动标注后，各类实体会以不同颜色的子块形式呈现，用户可以手动对实体进行校对和补充，如将「夏后」标注为「人」。（图 11）



图 10



图 11

如果用户需要对某个实体手动进行统一标注，可以点击「实体标注」页面工具栏中的「匹配标注」选项，将同一个实体标注为某一类型，如将「春秋」统一标注为「时间」，这样所有的「春秋」实体都会被标注为时间。（图 12）

如果用户需要对多个实体手动进行统一标注，可以点击「实体标注」页面工具栏中的「批量匹配标注」选项，上传「匹配模式库」（图 13）。「匹配模式库」即以 Excel 表格的形式规定多个实体的类型，既支持普通文本标注实体，也支持正则表达式标注实体，可参考「匹配模式库」样例。（图 14）



图 12

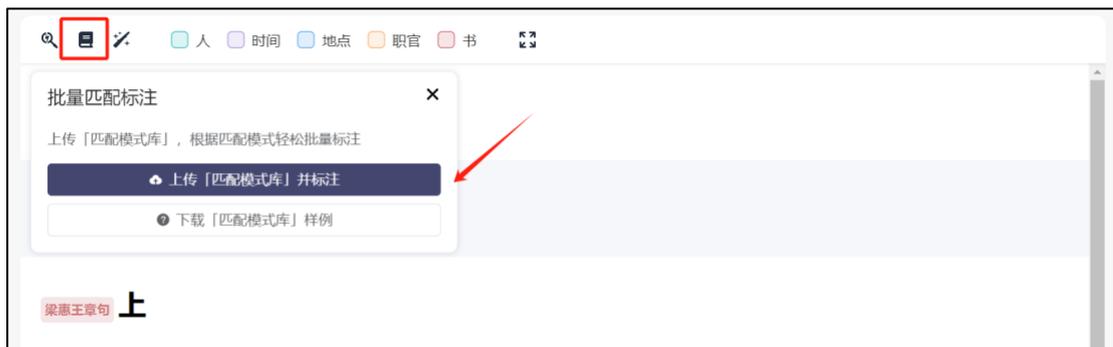


图 13

	A	B	C	D	E
1	pattern	flags	type	target	remarks
2	權		string	Person	普通文本匹配标注示例。Pattern是被匹配的文本，target是要标注的实体的ID。
3	劉?[備璋]		regex	Person	正则表达式标注示例。可匹配刘备、刘璋、备、璋。
4	tom	i	regex	Person	正则表达式标注示例。使用了正则表达式的i标记指明忽略大小写。
5					
6					
7					

图 14

用户还可以自己定义实体类型。在「实体标注」页面左上角的「实体类」一栏中点击「上传定义」，定制满足个性化标注需求的实体类型（图 15）。定义实体类型请参考「实体样例」。（图 16）



图 15

	A	B	C	D
1	id	text	description	path
2	Person	人	实体 - 人名	/
3	Time	时间	实体 - 时间	/
4	Location	地点	实体 - 地名	/
5	Position	职官	实体 - 职官	/
6	Book	书	实体 - 书名	/
7				
8				
9				

图 16

识别出来的实体会在左侧下方显示，点击实体则可以锁定文本中包含此实体的段落，实现窗格间的联动和内容重组。（图 17）

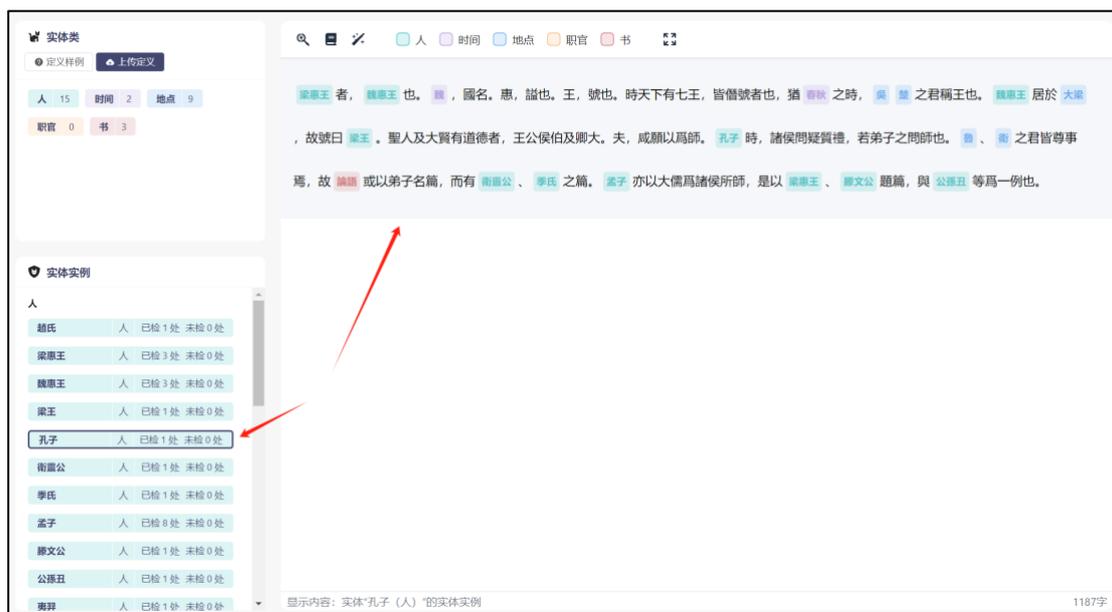


图 17

3.5. 关系标注

在「关系标注」页面，用户可以对标注出来的实体进行自动关系抽取。点击「关系标注」页面工具栏上的「自动标注关系」，文档会根据语义为两个实体标注关系。（图 18）



图 18

如果用户要手动定义实体间的关系，可以点击需要标注的实体，如先后点击「孟子」和「梁」（图 19），再选择「孟子，到达，梁」的选项（图 20），即可在二者间建立关系。



图 19



图 20

识别出来的关系实例会在左侧下方显示，点击某一关系实例则可以锁定文本中包含此关系的段落，实现窗格间的联动和内容重组。（图 21）



图 21

3.6. 生成知识图谱

如果用户需要将标注的实体及其关系以可视化的方式呈现出来，则可以点击「知识图谱」页面，即可在下方看到诸实体及其关系的知识图谱形式。（图 22）

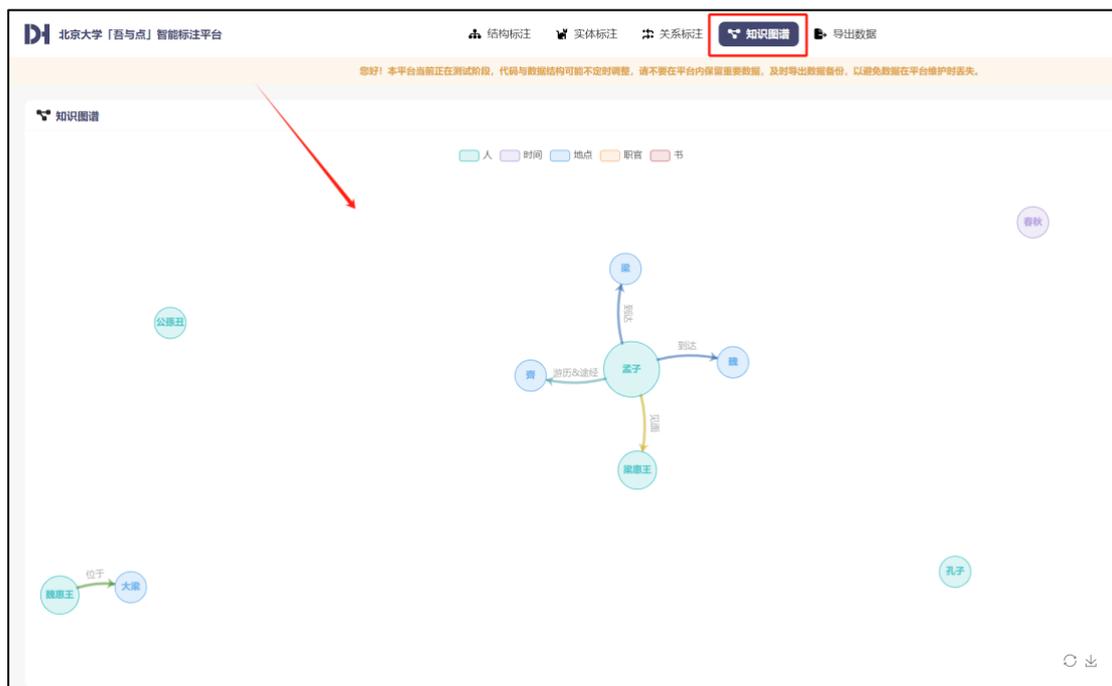


图 22

3.7. 导出数据

在所有文本数据都生成之后，用户可以将数据导出为特定的格式。点击「导出数据」的页面，可以看到各种数据导出的格式，包括「吾与点平台数据格式」、「序列标注数据格式」、「二维表格数据格式」和「其他工具数据格式」，用户点击相应的格式选项即可下载。（图 23）



图 23