WYD Help Center

1. WYD(Widen Your Data) Platform Introduction

WYD(Widen Your Data) Intelligent Annotation Platform is a general-purpose intelligent annotation platform developed by the Digital Humanities Research Center of Peking University. This platform supports intelligent structuring, entity annotation, relationship annotation, and automatic generation of knowledge graphs for free text, and exports various commonly used data formats such as Brat, Doccano, Gephi, Echarts, and BIOES. The platform integrates various deep learning AI models and achieves efficient human-machine collaboration through sophisticated interactive design. Currently, WYD has implemented entity recognition, relationship extraction, and graph generation on Classical Chinese texts, and entity recognition on Modern Chinese and English texts. In the future, it will support processing more languages.

2. The goals and solutions of the WYD platform

Goal: Build an intelligent data generation platform with excellent user experience, capable of supporting multiple users and multilingual environments. Complete a full range of tasks for intelligent data generation, handle complex and lengthy historical language data, and integrate with various other data analysis platforms.

Solution: Train AI models on a dataset of 3.3 billion characters of ancient Chinese language data to achieve automatic punctuation, automatic entity

recognition, and relationship extraction in language texts. Build an intelligent annotation platform to integrate project management, file organization, and data annotation functions, facilitating team collaboration.

3. WYD platform quick usage process

3.1. Upload text data

First, users upload their text data on the project management page. Upload the text data by following the path "Create Project (Figure 1) -> Open Project (Figure 2) -> Create Document/Import Document (Figure 3 and Figure 4) -> Open Document (Figure 5)".

If only a single document needs to be annotated, you can click the "Create Document" button, enter the document's name in the pop-up dialog box, and click the "Confirm" button. After that, you will see the empty document you created appear in the project's document list. (Figure 3).

If there are multiple texts that need to be uploaded in bulk, you can click the "Import Document" button (Figure 4). The "Wu Yu Dian" platform currently only supports importing text files in txt format and does not support importing Word or WPS files. After selecting the file, click the "Start Import" button. After the upload is complete, the page will automatically refresh, and the imported document will also appear in the project's document list.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

3.2. Structural Annotation

After opening the file, the first page you will enter is the "Structural Annotation" page. If you need to automatically add punctuation to the text, you can click the "Auto Punctuation" button (Figure 6). Periods, commas, and other punctuation marks will be automatically added to the text. If there are any errors in the punctuation, you can click the mouse to position the cursor at the incorrect location and manually edit or delete the punctuation (Figure 7).

Figure 6

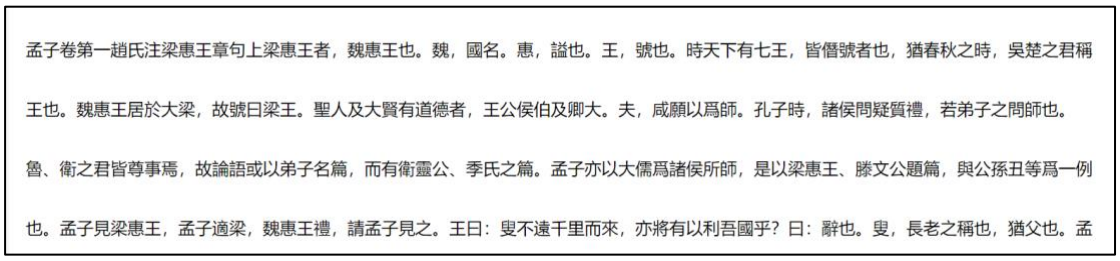

Figure 7

Users can also manually add text structures. On the "Structure Annotation" page, there is a text structure tool in the toolbar. Users can select specific text and annotate it as a certain structure type, such as setting "孟子卷第一" as a "卷" (Figure 8). The text structure will be displayed in the left column (Figure 9).

图 8

图 9

3.3. Entity Annotation

Then enter the "Entity Annotation" page. On this page, users can automatically annotate and extract entities from the text. Click on the "Auto Annotate Entities" function in the toolbar of the "Entity Annotation" page to automatically annotate the entities. (Figure 10) After automatic annotation,

different types of entities will be presented in the form of colored sub-blocks. Users can manually review and supplement the entities, such as annotating "夏后" as "人". (Figure 11)

Figure 10

Figure 11

If users need to manually annotate a specific entity, they can click on the "Entity Annotation" option in the toolbar of the "Entity Annotation" page, and annotate the same entity as a certain type, such as annotating "春秋" as "时间", so that all "春秋" entities will be annotated as time. (Figure 12)

If users need to manually annotate multiple entities, they can click on the "Batch Matching Annotation" option in the "Entity Annotation" page toolbar and upload the "Matching Pattern Database" (Figure 13). The "Matching Pattern Database" is an Excel spreadsheet that specifies the types of multiple entities,

supporting both regular text annotation entities and regular expression annotation entities. Please refer to the example in the "Matching Pattern Database" (Figure 14).

Figure 12

Figure 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	正则表达式标注实例。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					

Figure 14

Users can also define their own entity types. Click on "Upload Definition" in the "Entity Types" column at the top left corner of the "Entity Annotation" page to customize entity types that meet personalized annotation needs (Figure 15). Please refer to the "Entity Sample" for defining entity types (Figure 16).

Figure 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				

Figure 16

The recognized entities will be displayed at the bottom left, clicking on an entity will lock the paragraphs containing this entity in the text, achieving linkage between panels and content reorganization. (Figure 17)

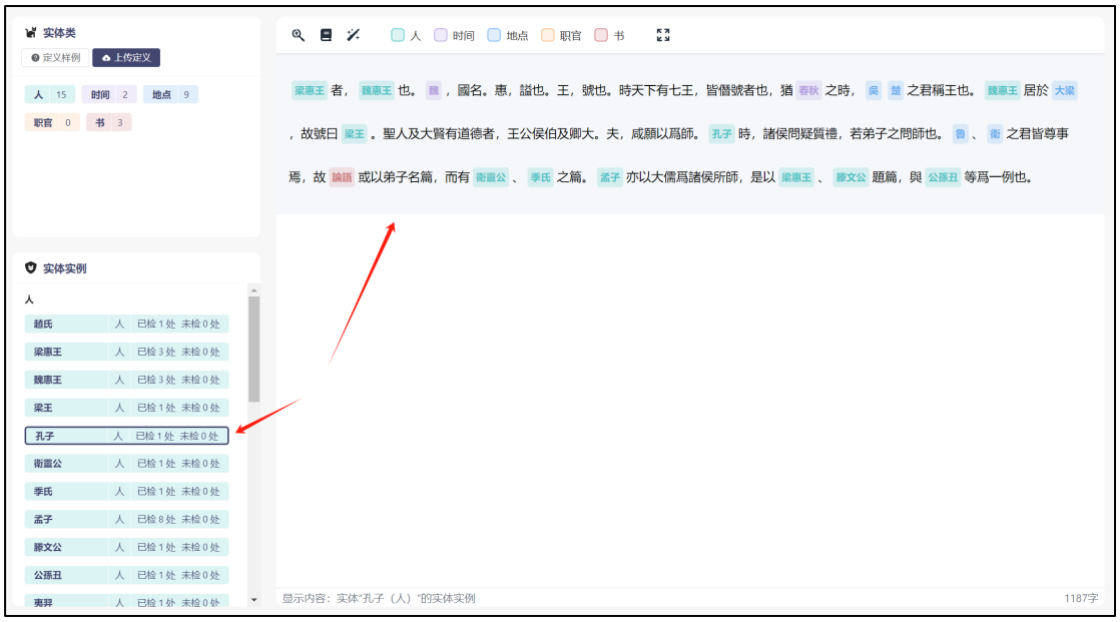

Figure 17

3.5. Relationship Annotation

On the "Relation Annotation" page, users can automatically extract relationships from annotated entities. By clicking on "Auto Annotate Relations" on the toolbar of the "Relation Annotation" page, the document will annotate the relationship between two entities based on semantics. (Figure 18)

Figure 18

If the user wants to manually define the relationship between entities, they can click on the entities that need to be annotated, such as clicking on "孟子" and "梁" in sequence (Figure 19), and then selecting the option "孟子，到达，梁" (Figure 20) to establish a relationship between them.

Figure 19

Figure 20

The recognized relationship instances will be displayed on the bottom left. Clicking on a relationship instance will lock the paragraphs containing this relationship in the text, achieving linkage between panels and content reorganization. (Figure 21)

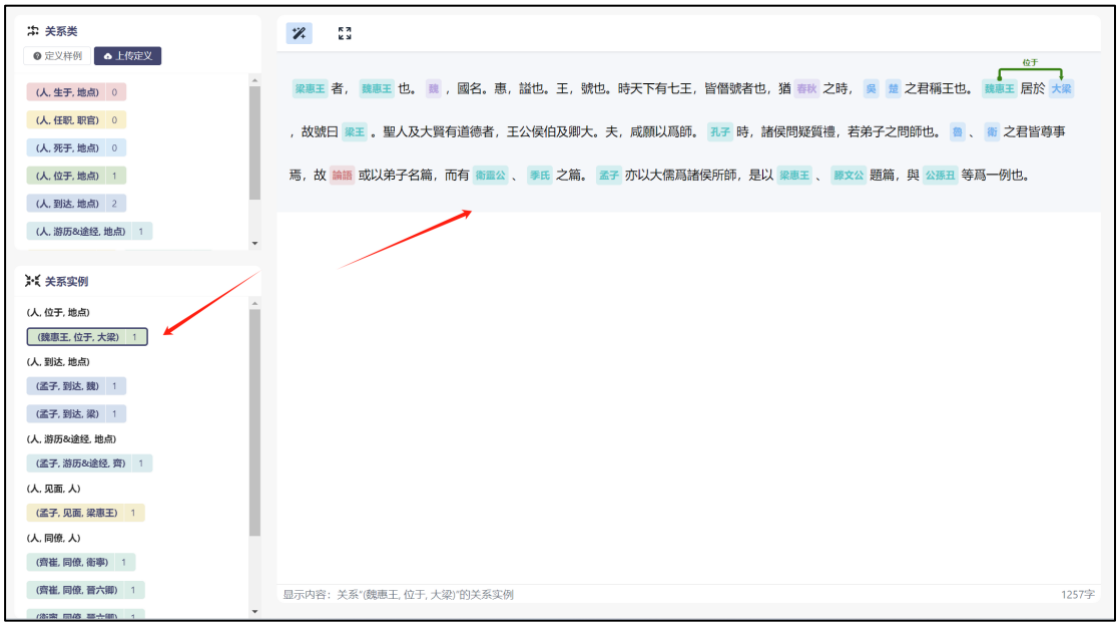

Figure 21

3.6. Generate Knowledge Graph

If users need to visualize the annotated entities and their relationships, they can click on the "Knowledge Graph" page to see the knowledge graph representation of various entities and their relationships below. (Figure 22)

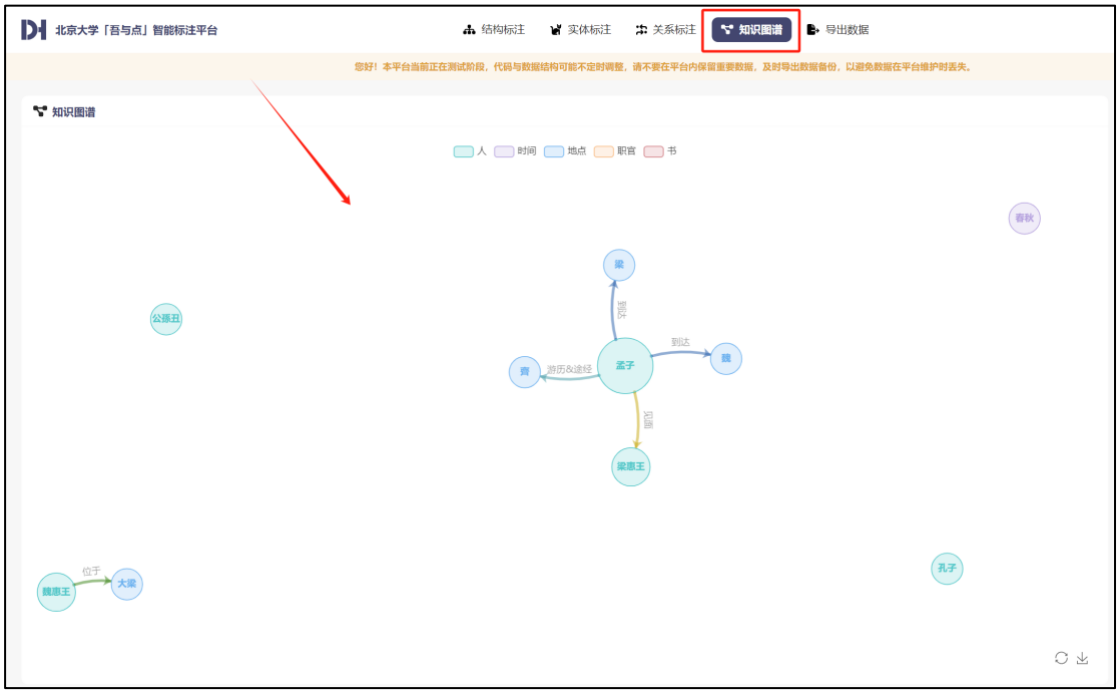

Figure 22

3.7. Export Data

After all the text data has been generated, users can export the data in specific formats. On the "Export Data" page, various data export formats are available, including "WYD Platform Data Format," "Sequence Labeling Data Format," "Two-dimensional Table Data Format," and "Other Tool Data Format." Users can click on the corresponding format option to download. (Figure 23)

Figure 23