

本项目得到国家科技创新2030“新一代人工智能”重大项目——“以自然语言为核心的语义理解理论、模型与方法”（项目号：2020AAA0106701）支持。

SpaCE2022 标注

—— task3 空间信息标注

<https://2030nlp.github.io/Sp22AnnoOL/menu>

詹卫东

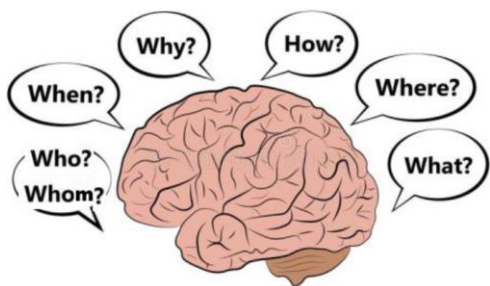
北京大学中文系

zwd@pku.edu.cn

中文空间语义理解评测任务 (**S**patial **C**ognition **E**valuation, 简称SpaCE)

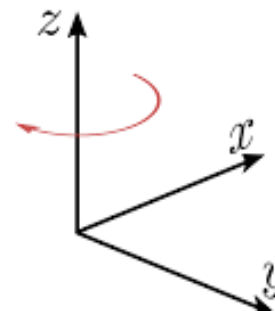


中文空间语义理解评测任务 (Spatial Cognition Evaluation, 简称SpaCE)



人类

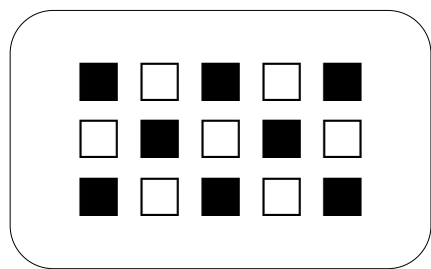
空间
认知



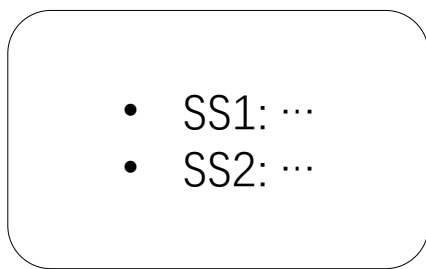
计算机



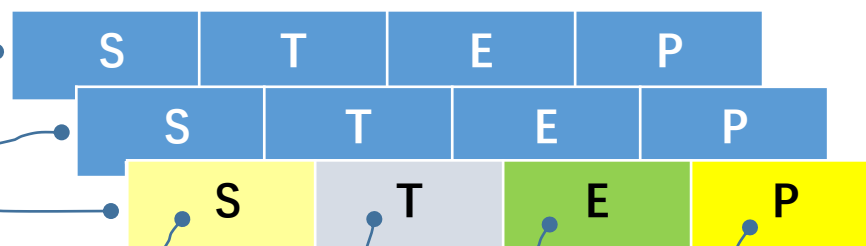
一段文本



几个句子



STEP 空间信息标注



S: 空间实体

P: 空间方位

E: 事件

T: 时间

某物（或某人）+ 某时 + 因 某事 + 在 某处

某物（或某人）+ 某时 + 为 某事 + 在 某处

某物（或某人）+ 某时 + 因 某事 + 往 某处

某物（或某人）+ 某时 + 为 某事 + 从 某处 往 某处

某物（或某人）+ 某时 + 做 某事 + 经 某处 往 某处

.....

string

string

arg0

string

处所

...

arg1

string

部件处所

...

arg2

string

部位

...

说话时

E开始前

E结束后

E时间

长时

arg3

string

起点

...

argM

string

终点

...

argS

string

方向

...

argT

string

朝向

...

路径

...

形状

...

距离

...

非结构化空间信息



结构化空间信息

汤姆平常放在床头的泰迪熊不见了

S	T	E	P	F
泰迪熊	平常	放	在床头	
泰迪熊	说话时	不见	在床头	f

E	放
arg0	汤姆
arg1	泰迪熊
argS	在床头
argT	平时

E	不见
arg0	泰迪熊
arg1	
argS	
argT	

村边，一条干枯的河床上，几个新堆起的小土包一字排开。阿曾和几个刑事犯一起被押到土包前。她身后的两个汉子狠劲地把她摠跪在地上。

- SS1： 村边，一条干枯的河床上，
几个新堆起的小土包一字排开
- SS2： 阿曾和几个刑事犯一起被押到土包前
- SS3： 她身后的两个汉子狠劲地把她摠跪在地上

S	T	E	P	F
河			处所: 村边	
几个小土包			处所: 河床上，形状: 一字排开	
阿曾和几个刑事犯		押	终点: 到土包前	
两个汉子			处所: 她身后	
她		摠	终点: 在地上	
她		跪	处所: 在地上	

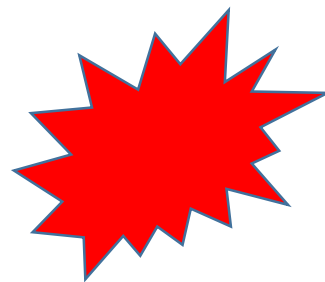
E	arg0	arg1	arg2	arg3	argM	argS	argT
押		阿曾和几个刑事犯				到土包前	
摠	两个汉子	她				在地上	
跪	她					在地上	

同指关系		
她	=	阿曾

标注工作的“尖锐”矛盾

信息量：越充分越好

标注量：越简单越好



NLP不断增长的数据信息量需求，跟一般人“**能不动手就不动手**”的美好愿望之间的激烈冲突

信息标注需求：

S 注意事项

P 注意事项

E 注意事项

T 注意事项

R 注意事项

F 注意事项

- (1) 什么实体是S? —— 有P信息的实体;
- (2) 怎么表示S? —— 能区别其他S, 能跟P信息对应;
- (3) 哪些空间算P信息? —— 物理空间、信息空间、认知空间
- (4) 同指实体标几个S? —— 标一个就行
- (5) 并列结构标几个S? —— 同P则合, 异P则分

信息标注需求：

S 注意事项

P 注意事项

E 注意事项

T 注意事项

R 注意事项

F 注意事项

序号	特征	特征值	含义
1.	处所	处所词 处所词+方位词 介词+处所词+方位词	处所，指一个静态空间实体相对某外部参照物的位置。 例如：泰迪熊 在沙发上
2.	部件处所	处所词 处所词+方位词 介词+处所词+方位词	部件处所，指作为一个部件的空间实体在整体中的位置。 例如：鼻子 在脸的中间
3.	部位	处所词 处所词+方位词 介词+处所词+方位词	部位，指以空间实体为参照，描述其内部部件的位置。 例如：汽车 右侧 、小客车 左前部
4.	起点	处所词 处所词+方位词 介词+处所词+方位词	起点，指在动态空间实体的方位发生变化的场景下，变化开始时实体的处所。 例如： 从北大东门 出发
5.	终点	处所词 处所词+方位词 介词+处所词+方位词	终点，指在动态空间实体的方位发生变化的场景下，变化结束时实体的处所。 例如： 送到医院
6.	方向	介词+方位词 处所词+介词+方位词 处所词、方位词、趋向动词	方向，指动态空间实体位移方向。 空间实体在动态中才有“方向”特征。 例如：车 往南 开、他 爬上去 了、 外 推2米
7.	朝向	介词+方位词 介词+处所词	朝向，指空间实体的不同侧面的位置。 例如：床头 朝东 、背 对着镜子
8.	路径	形状值 处所词 <处所词 ₁ ，处所词 ₂ ，...>	路径，指空间实体位移时经过的轨迹。 例如：京沪线经过 蚌埠 和 无锡
9.	形状	直线 圆 圆圈儿 弯 半圆	形状，指（1）一个空间实体自身的形状； （2）若干个空间实体组成的形状。
10.	距离	数词+单位词 远 近 变远 变近	距离，指两个空间实体看作两个点，度量两点连线的长度。

信息标注需求：

S 注意事项

P 注意事项

E 注意事项

T 注意事项

R 注意事项

F 注意事项

序号	角色标记	对应论旨角色	含义
1.	arg0	施事（共同施事）、当事（间接当事）	核心角色1：主体
2.	arg1	受事、系事、结果、对象、内容	核心角色2：客体
3.	arg2	与事	核心角色3：受益/受损者
4.	arg3	工具、材料、方式	外围角色1：伴随者（准外围）
5.	argM	事量、范围、原因、目的	外围角色2：环境成分（真外围）
6.	argS	起点、终点、路径、方向、处所	外围角色3：空间角色
7.	argT	起始、结束、时点、时段	外围角色4：时间角色

信息标注需求：

S 注意事项

P 注意事项

E 注意事项

T 注意事项

R 注意事项

F 注意事项

序号	类型	时间字段	时间值	示例 / 说明
1.	时值时间	原文时间	时间义词语	上午、13: 15, 2022年6月25日, 儿童节, 夏至
2.	说话时间	参照时间	说话时	参考语境说话人时间信息, “说话时”为系统关键字
3.	说话时间	参照时间	过去	参考语境说话人时间信息, “过去”为系统关键字
4.	说话时间	参照时间	将来	参考语境说话人时间信息, “将来”为系统关键字
5.	事件时间	参照事件+参照时间	V之前	参考事件V的时间信息, V是句中动词
6.	事件时间	参照事件+参照时间	V之后	参考事件V的时间信息, V是句中动词
7.	事件时间	参照事件+参照时间	V之时	参考事件V的时间信息, V是句中动词
8.	事件时间	参照事件+参照时间	V之间	参考事件V的时间信息, V是句中动词

信息标注需求：

S 注意事项

P 注意事项

E 注意事项

T 注意事项

R 注意事项

F 注意事项

- (1) 同指关系单独标注，标注界面有专门的标注区；
- (2) 只需标注有P信息的空间实体的同指关系；
- (3) 有同指关系的空间实体，共享STEP信息；
- (4) 同指关系标注，以指代词和指称词的同指为主；
- (5) 其他名词组的同指关系，如能减少STEP标注量，也可标注

- (1) F (Facticity) 信息一般不需要标注，表示默认值为t（真）；
- (2) 如果标注F信息，程序会自动把F信息的值设置为f（假）；

信息标注需求：

8 E 8 T 10 P

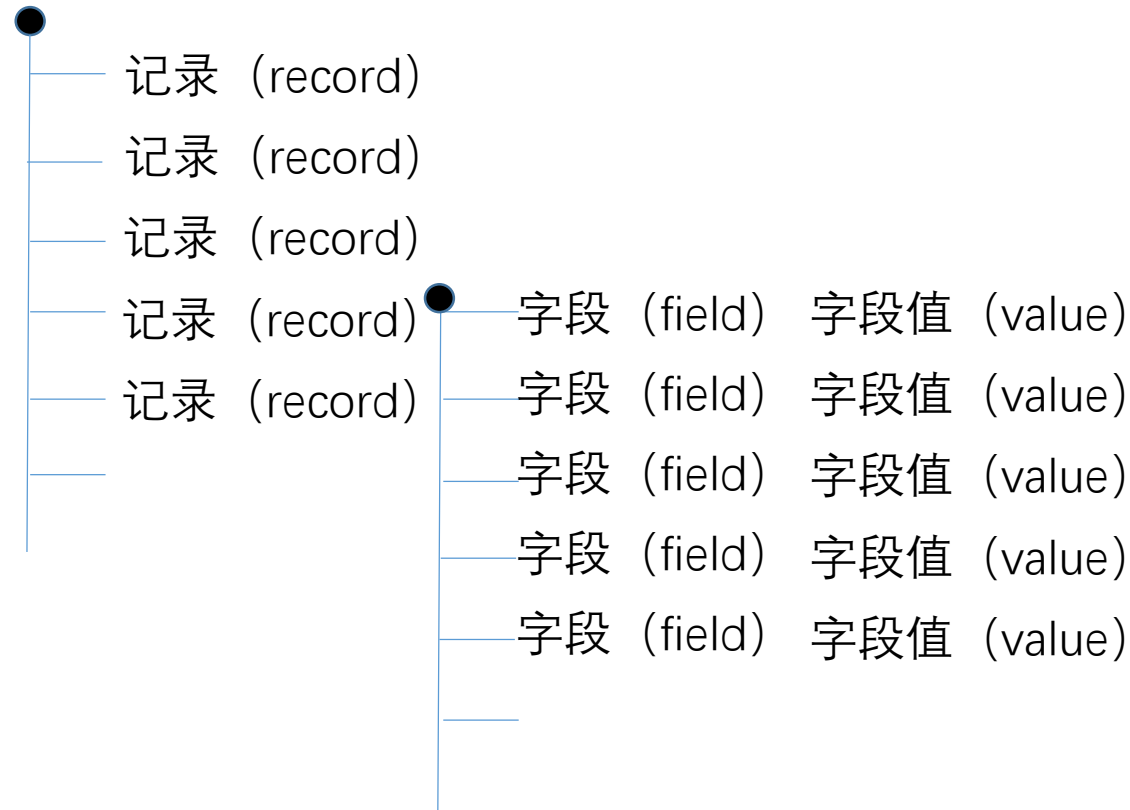
信息大项	大项数量	信息小项	小项数量	合计
S	1		0	1
E	1	arg0, arg1, arg2, arg3, argM, argS, argT	7	8
T	0	"", 说话时, 过去, 将来, 之前, 之后, 之间, 之时	8	8
P	0	处所, 部件处所, 部位, 起点, 终点, 方向, 朝向, 路径, 形状, 距离	10	10
F	1		0	1
R	1		0	1
合计	4		25	29

“特殊”成分（形式）处理方法

1. 并列成分（并置标注）
2. 修饰成分（区分主次）
3. 分离成分（拼接合体）
4. 同指成分（共享信息）
5. 隐性信息（点到为止）

数据层级

数据 (data)



记录

S-T-E-P 整体信息标注 [2]

克隆 删除

字段	实体S	拷	贴	字段值	【请在文中选取】	存	删
字段	处所	拷	贴	字段值	【请在文中选取】	存	删
字段	起点	拷	贴	字段值	【请在文中选取】	存	删
字段	终点	拷	贴	字段值	【请在文中选取】	存	删
字段	方向	拷	贴	字段值	【请在文中选取】	存	删
字段	朝向	拷	贴	字段值	【请在文中选取】	存	删
字段	事件E	拷	贴	字段值	【请在文中选取】	存	删
字段	原文时间	拷	贴	字段值	【请在文中选取】	存	删
字段	参照事件	拷	贴	字段值	【请在文中选取】	存	删
字段	参照时间	拷	贴	字段值	【请选择】	存	删

添加字段

<请选择>

操作注意事项

- 拷贝
- 粘贴
- 修改
- 删除 ×
- 存储
- 增添
- 拼接 ⊕
- 并置 填入
- 关闭

- 点“删除”（在“克隆”的旁边）：表示删除当前记录（record），即删除整个信息标注区；
- 点“删”：表示删除当前字段（field），即删除字段名称及其同一行文本框；
- 点 × 号：表示删除当前值（value），即删除一个文本片段。
- 有点意外 😊：点击3个信息标注区右上角的 × 号，表示关闭当前标注区，不是删除。

新增记录

新增 S-T-E-P 整体信息标注

新增 事件角色信息标注

新增 同指关系标注

新增字段

添加字段

<请选择>

+

克隆记录

🔗 S-T-E-P 整体信息标注 [10]

克隆 删除

÷ ×

操作注意事项

- 拷贝
- 粘贴
- 修改
- 删除 ×
- 存储
- 增添
- 拼接 ⊕
- 并置 填入
- 关闭

正在标注

事件角色信息标注 [0] 克隆 删除

事件谓词	拷	贴			
			【请在文中选取】	存	删
arg0			【请在文中选取】	存	删
arg1			【请在文中选取】	存	删
arg2			【请在文中选取】	存	删
arg3			【请在文中选取】	存	删
argM			【请在文中选取】	存	删
argS			【请在文中选取】	存	删
argT			【请在文中选取】	存	删

暂时保存 完成并保存 重置 清空 DEBUG

上一条 下一条

Q & A

zwd@pku.edu.cn

<https://2030nlp.github.io/Sp22AnnoOL/menu>