

# 知识图谱准确率和prompt亮点

## prompt工程

### Prompt的构成四要素

- 指示 (Instructions) - 关键词：任务描述
  - 指示是对任务的明确描述，相当于给模型下达了一个命令或请求。它告诉模型应该做什么，是任务执行的基础。
- 上下文 (Context) - 关键词：背景信息
  - 上下文是与任务相关的背景信息，它有助于模型更好地理解当前任务所处的环境或情境。在多轮交互中，上下文尤其重要，因为它提供了对话的连贯性和历史信息。
- 输入 (Input) - 关键词：数据输入
  - 输入是任务的具体数据或信息，它是模型需要处理的内容。在Prompt中，输入应该被清晰地标识出来，以便模型能够准确地识别和处理。
- 输出 (Output) - 关键词：结果格式
  - 输出是模型根据输入和指示生成的结果。在Prompt中，通常会描述输出的格式，以便后续模块能够自动解析模型的输出结果。常见的输出格式包括结构化数据格式如JSON、XML等。

### Few-shot提示学习法

通过输入一些类似问题和问题答案，让模型参考学习，并在同一个prompt的末尾提出新的问题，以此提升模型的推理能力。

## 实体识别和关系抽取

[https://github.com/zjunlp/DeepKE/blob/main/README\\_CN\\_SCHEMA\\_CN.md](https://github.com/zjunlp/DeepKE/blob/main/README_CN_SCHEMA_CN.md)

实体类型和关系类型参考浙江大学开源中文知识图谱DeepKE的定义。

reference (知识图谱的论文，贴ppt上显得专业)：arXiv:2201.03335 [cs.CL]

## prompt

- 1 实体：在知识图谱中，实体代表了现实世界中的个体或者概念。关系用来描述实体之间的连接或联系，它表示了实体之间的某种关联关系。下面给出的关系类型中比如("地点", "人物", "祖籍")，第一个元素表示"地点"表示宾语，第二个元素"人物"表示主语，第三个元素"祖籍"表示关系类型，所以这个三元组表示人物的祖籍是地点。对于给定的文本，需要从中识别出实体和它们之间的关系，你输出的实体

类型是下面给出的实体类型(可以有别的类型), 输出的关系类型是下面给出的关系类型(可以有别的类型), 输出json格式的字符串, 你必须按照示例的格式输出。

- 2 实体类型: 人物, 影视作品, 生物, Number, Date, 国家, 网站, 网络小说, 图书作品, 歌曲, 地点, 气候, 行政区, TEXT, 历史人物, 学校, 企业, 出版社, 书籍, 音乐专辑, 城市, 景点, 电视综艺, 机构, 作品, 语言, 学科专业
- 3 关系类型: ("地点", "人物", "祖籍"), ("人物", "人物", "父亲"), ("地点", "企业", "总部地点"), ("地点", "人物", "出生地"), ("Number", "行政区", "面积"), ("Text", "机构", "简称"), ("Date", "影视作品", "上映时间"), ("人物", "人物", "妻子"), ("音乐专辑", "歌曲", "所属专辑"), ("Number", "企业", "注册资本"), ("城市", "国家", "首都"), ("人物", "影视作品", "导演"), ("Text", "历史人物", "字"), ("Number", "人物", "身高"), ("企业", "影视作品", "出品公司"), ("Number", "学科专业", "修业年限")
- 4 input: 合肥土地市场迎来龙年首场土拍, 共计3宗涉宅地块出让, 地块均溢价成交, 最高溢价率达54%, 共计成交金额32.3亿元。
- 5 output: {"entities": [{"type": "地点", "value": "合肥"}, {"type": "Date", "value": "龙年"}, {"type": "Number", "value": "3宗"}, {"type": "Text", "value": "涉宅地块"}, {"type": "Number", "value": "54%"}, {"type": "Number", "value": "32.3亿元"}], "relations": [{"first": {"type": "地点", "value": "合肥"}, "second": {"type": "Date", "value": "龙年"}, "relation": "时间地点"}, {"first": {"type": "Number", "value": "3宗"}, "second": {"type": "Text", "value": "涉宅地块"}, "relation": "数量类型"}, {"first": {"type": "Number", "value": "54%"}, "second": {"type": "Text", "value": "涉宅地块"}, "relation": "属性数值"}, {"first": {"type": "Number", "value": "32.3亿元"}, "second": {"type": "Text", "value": "涉宅地块"}, "relation": "价格类型"}]}
- 6 input:

## prompt构成四要素

- 指示: 对于给定的新闻文本, 需要从中识别出实体和它们之间的关系
- 上下文: 实体类型、关系类型
- 输入: 需要进行实体识别和关系抽取的新闻文本
- 输出: 你输出的实体类型是下面给出的实体类型(可以有别的类型), 输出的关系类型是下面给出的关系类型(可以有别的类型), 输出json格式的字符串, 你必须按照示例的格式输出。

## Few-shot提示学习法

在以上的prompt中给出了一个示例问题和答案让LLM进行参考学习, 并在同一个prompt的末尾提出新的问题

## 事件抽取

### Prompt

- 1 信息抽取中事件的定义：事件是发生在某个特定的时间段、某个特定的地域范围内，由一个或者多个角色参与的一个或者多个动作组成的事件或者状态的改变。事件类型：不同动作或者状态的改变代表不同类型的事件。事件元素：同一类型的事件中不同的时间、地点、角色代表了不同的事件实例。你需要提取出事件的类型和事件发生的时间，并按照时序关系（即事件发生的时间顺序升序）输出，未给出具体时间的事件不要输出，下面的示例中一个输入可能会有多个事件，按照示例的格式输出。
- 2 input: 1992年10月3日，奥巴马与米歇尔在三一联合基督教堂结婚。
- 3 output: [{"event": "结婚", "time": "1992年10月3日"}]
- 4 input: 《每日经济新闻》获悉，近日英国富时罗素指数公司公布了富时全球股票指数半年度指数审查报告。根据报告，富时全球股票指数系列将新调入A股76只、调出1只。据悉，此次被调入的大盘股共有10只，包括首创证券、石药集团、国博电子、海光信息、建元信托、华大智造、荣昌生物、联影医疗、江波龙、唯捷创芯等；微盘股有2只，分别为容知日新和科净源；同时还调入了数十只小盘股。根据近日市值计算，A股占富时罗素新兴市场指数市值将从约5.71%增加至约6.18%，占富时全球指数市值从约0.55%增加至约0.61%。调整措施将于3月4日生效，3月15日进行指数调仓。
- 5 output: [{"event": "调整措施将于3月4日生效", "time": "3月4日"}, {"event": "3月15日进行指数调仓", "time": "3月15日"}]
- 6 input: 来源：格隆汇格隆汇2月29日 | 金花股份(600080.SH)公布，截止2024年2月29日，公司通过集中竞价交易方式累计回购公司股份1123.29万股，已回购股份占公司总股本的3.00%，与上次披露数相比增加0.48%，回购成交的最高价为7.79元/股，最低价为6.90元/股，已支付的资金总额为人民币8386.34万元（不含印花税、交易佣金等交易费用）。
- 7 output: [{"event": "公司回购股份", "time": "2024年2月29日"}, {"event": "回购股份占公司总股本的3.00%", "time": "2024年2月29日"}, {"event": "回购成交的最高价为7.79元/股", "time": "2024年2月29日"}, {"event": "最低价为6.90元/股", "time": "2024年2月29日"}, {"event": "已支付的资金总额为人民币8386.34万元", "time": "2024年2月29日"}]
- 8 input:

## prompt构成四要素

- 指示：你需要提取出事件的类型和事件发生的时间
- 上下文：信息抽取中事件的定义：事件是发生在某个特定的时间段、某个特定的地域范围内，由一个或者多个角色参与的一个或者多个动作组成的事件或者状态的改变。事件类型：不同动作或者状态的改变代表不同类型的事件。事件元素：同一类型的事件中不同的时间、地点、角色代表了不同的事件实例。
- 输入：需要进行事件抽取的新闻文本
- 输出：并按照时序关系（即事件发生的时间顺序升序）输出，未给出具体时间的事件不要输出，下面的示例中一个输入可能会有多个事件，按照示例的格式输出。

## Few-shot提示学习法

在以上的prompt中给出了几个示例问题和答案让LLM进行参考学习，并在同一个prompt的末尾提出新的问题