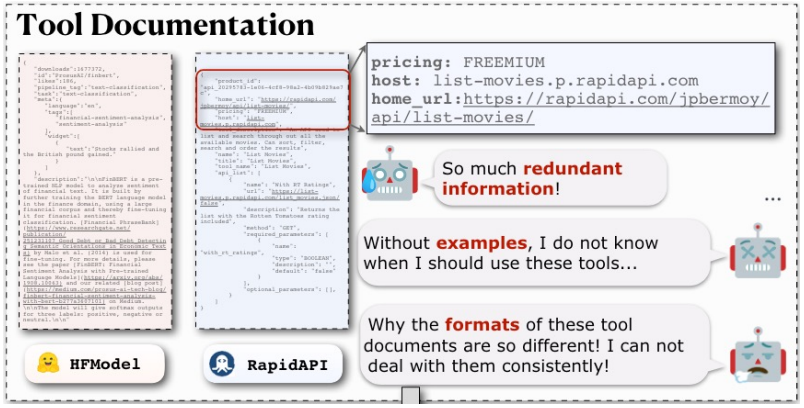
EasyTool (arXiv 202401)

Motivation: 现有基于LLM的工具学习主要使用工具文档理解工具，但是目前的文档可能是多种多样、冗余或者不完整的，这极大地影响了LLM使用工具的能力。
整体思路：提出EasyTool，将多样化、冗长的工具文档转换成统一、简洁的工具指令，为基于LLM的Agent提供标准化的工具表述和功能，以方便工具使用。

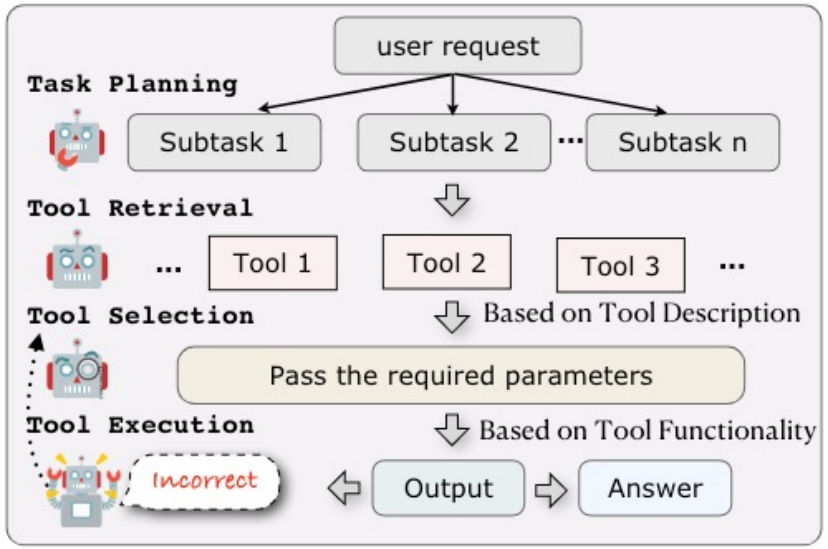
现有工具学习的两种路线

- 微调开源LLM以生成使用工具的函数调用。该方法需要构造额外的数据集和工具库用于训练，**扩展性较差**；无法以即插即用的方式集成外部工具，**灵活性较差**。
- 通过提供工具文档和工具功能的少量示例来检索和调用外部工具。这种方式带来了**巨大的Prompt消耗**，阻碍了基于LLM的Agent发展。



工具文档包含工具表述、工具参数、示例等信息，是LLM理解和使用工具不可或缺的组成部分。但目前基于文档的工具调用还存在以下问题(如上图所示)：
不一致性：不同的工具具有不一致且多样化的文档格式；
冗余：文档中的冗余信息，导致Prompt中token过度消耗；
不完整：有时工具文档会缺乏关键信息，影响工具的利用。

如何解析文档、提取最本质的信息并提供统一的格式是十分重要的。



Pipeline分为**四个阶段**：
Tool Planning: 分析用户请求并将其拆分为特定的子任务；
Tool Retrieval: 为每个子任务检索相似度最高的K个工具；
Tool Selection: 从检索工具集中选择最合适的工具；
Tool Execution: 执行工具，如果执行失败则回退到工具选择阶段重新选择工具，直至执行成功或到达最大尝试次数。

I: Tool Description Generation

```
/* I: Task prompt */
Your task is to create a concise and effective tool usage description based on the tool documentation. You should ensure the description only contains the purposes of the tool without irrelevant information. Here is an example:
/* Examples */
{Tool Documentation}
Tool usage description:
{Tool_name} is a tool that can {General_Purposes}.
This tool has {Number} multiple built-in functions:
1. {Function_1} is to {Functionality_of_Function_1}
2. {Function_2} is to ...
/* Auto generation of tool description */
{Tool Documentation of 'Aviation Weather Center'}
Tool usage description:
'Aviation Weather Center' is a tool which can provide official aviation weather data...
```

II: Tool Function Guidelines Construction

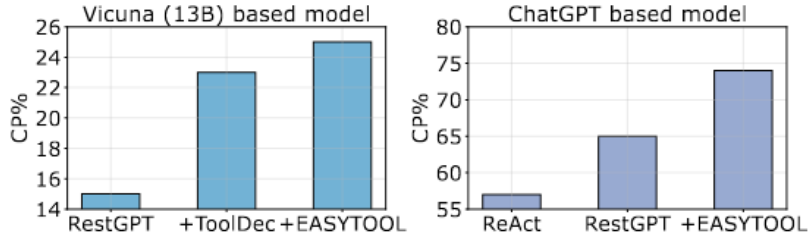
```
/* Task prompt */
Your task is to create the scenario that will use the tool.
1. You are given a tool with its purpose and its parameters list. The scenario should adopt the parameters in the list.
2. If the parameters and parameters are both null, you should set: {"Scenario": XX, "Parameters":{}}.
Here is an example:
/* Examples */
{Tool_name} is a tool that can {General_Purposes}.
{Function_i} is to {Functionality_of_Function_i}
{Parameter List of Function_i}
One scenario for {Function_i} of {Tool_name} is:
{"Scenario": XX, "Parameters":{XX:XX}}
/* Auto-construction for Tool Function Guidelines */
'Ebay' can get products from Ebay in a specific country.
'Product Details' in 'Ebay' can get the product details for a given product id and a specific country.
{Parameter List of 'Product Details'}
One scenario for 'Product Details' of 'Ebay' is:
{"Scenario": "if you want to know the details of the product with product ID 1954 in Germany from Ebay",
"Parameters":{"product_id": 1954, "country": "Germany"}}.
```

1. 整理原始工具文档，剔除不相关信息，保留功能描述；
2. 设计功能指南指令，通过提供工具的必需和可选参数，以及LLM示例使LLM可以进一步完善工具文档。

Model	Method	I2-Category			I3-Instruction			Average		
		Pass	Win	Success	Pass	Win	Success	Pass	Win	Success
ChatGPT	ReACT	39.0	-	18.0	23.0	-	1.0	31.0	-	9.5
	DFSDT	64.5	63.0	24.0	60.0	70.0	6.0	62.3	66.5	15.0
	DFSDT-EASYTOOL	74.5	76.5	68.5	65.0	88.0	37.0	69.8	82.3	52.8
	DFSDT-EASYTOOL-Retriever	69.0	71.0	60.5	66.0	89.0	42.0	67.5	80.0	51.3
ToolLLaMA-7B	ReACT	30.0	45.5	9.5	22.0	49.0	3.0	26.0	47.3	6.3
	TOOLDEC	42.0	55.0	15.0	34.0	60.0	22.0	38.0	57.5	18.5
	DFSDT	66.0	55.0	24.0	56.0	56.0	6.0	61.0	55.5	15.0
	DFSDT-Retriever	57.0	60.0	11.5	54.0	69.0	2.0	55.5	64.5	6.8
Vicuna-7B	ReACT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	DFSDT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	DFSDT-EASYTOOL	72.5	77.0	40.5	68.0	81.0	34.0	70.3	79.0	37.3
	DFSDT-EASYTOOL-Retriever	75.0	68.0	46.5	67.0	85.0	36.0	71.0	76.5	41.3
Mistral-Instruct-7B	ReACT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	DFSDT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	DFSDT-EASYTOOL	75.0	76.0	56.0	66.0	87.0	38.0	70.5	81.5	47.0
	DFSDT-EASYTOOL-Retriever	74.5	71.5	54.5	68.0	88.0	38.0	71.3	79.8	46.3
GPT-4	ReACT	67.5	53.5	27.0	40.0	71.0	4.0	53.8	62.3	15.5
	DFSDT	69.5	57.0	42.0	59.0	73.0	50.0	64.3	65.0	46.0
	DFSDT-EASYTOOL	**76.5**	**78.5**	**76.0**	**69.0**	89.0	**64.0**	**72.8**	**83.8**	**70.0**
	DFSDT-EASYTOOL-Retriever	72.5	72.0	73.5	**69.0**	**90.0**	53.0	70.8	81.0	63.3

Real-World Question Answering

ToolBench: I2-Category(200), I3-Instruction(100)
通过EasyTool，可以让没有工具调用指令微调的模型也具备工具调用能力，如Vicuna和Minstral.



Real-World Web Services

RestBench (TMDB)
EasyTool可以帮助找到正确的工具解决路径

Model	One-hop (↑)	Multi-hop (↑)	Error (↓)
Vicuna-30B	15.00	1.00	-
+ CoT	13.33	4.00	-
+ ReAct	45.00	7.35	20.31
+ EASYTOOL	**65.00**	**11.76**	**10.15**
ChatGPT	55.00	9.00	-
+ CoT	48.33	17.64	-
+ ReAct	85.00	41.17	9.38
+ EASYTOOL	**91.66**	**48.53**	**2.34**

FuncQA:
68 one-hop
60 multi-hop

Numerical Reasoning

EasyTool可以推广到没有工具文档的工具调用