# **业务流程建模系统**

小组成员及分工：

卢毅榕：编写数据集爬虫，微调大模型

娄骁扬 ：ssn建模

何瀛龙： 写文档

杨阳：数据集筛选

# 摘要

本文设计了一种大模型辅助的物联网业务流程建模系统，旨在提高物联网应用中业务流程建模的效率和准确性。通过结合大模型的自然语言处理能力和物联网资源的复杂结构，我们提出了一种新的建模方法。该方法包括以下几个关键步骤：

首先，我们会建立物联网资源模型。物联网资源模型是整个系统的基础，其构建需要对物联网设备、属性及其相互关系进行详细描述。本文中，使用 OWL（Web Ontology Language）作为描述语言，因其强大的语义表达能力，能够精确地描述资源类型、实例以及属性关系。具体步骤包括：搜集领域知识、设计资源模板、创建资源实例、绑定数据属性、设置关系属性等。以光纤授时系统为例，我们详细描述了如何通过这些步骤构建物联网资源模型，并使用工具 Protégé 进行具体实现。

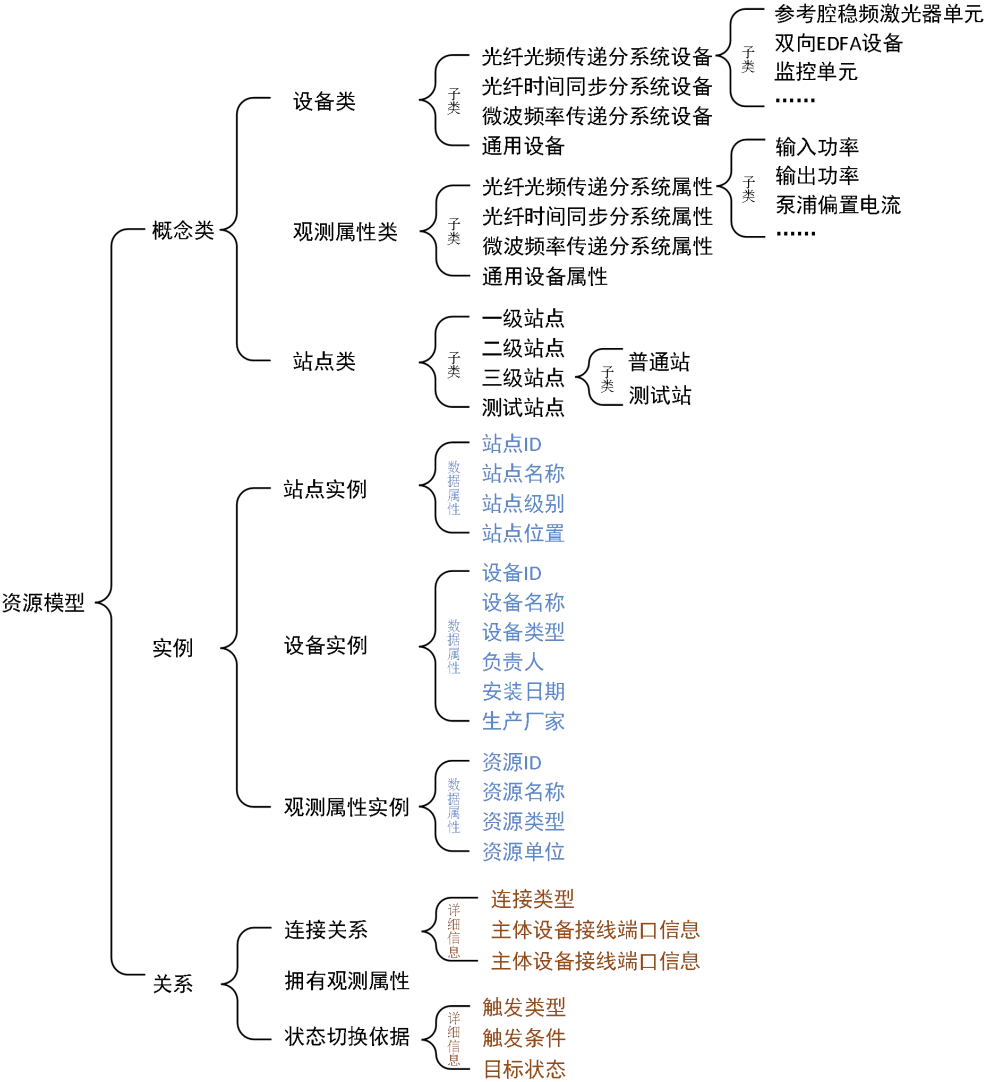
接下来，根据建立的资源模型，我们设计业务流程图。不同于命令式流程图，本文采用了声明式流程图的方法，特别是 DCR 图（Dynamic Condition Response Graphs），用以描述业务流程的逻辑结构和约束条件。DCR 图能够灵活地表达业务规则，动态调整流程路径，适用于复杂多变的物联网应用场景。我们详细介绍了 DCR 图的基本元素、构建方法以及在业务流程中的应用。

最后，借助训练的大模型，我们将业务流程图转换为自然语言描述。选择的 LLAMA（Language Learning Assisted by Machine and Algorithms）模型，基于深度学习和自然语言处理技术，通过在大规模文本数据上的预训练，具备了强大的语言理解和生成能力。本文使用飞桨平台的 PaddlePaddle 深度学习框架进行模型实现，并设计了数据爬虫和转换工具，生成标准化的训练数据集。模型训练过程中，我们详细设置了各种参数，确保模型能够准确地将业务流程图转换为自然语言描述。

通过上述步骤，本文实现了从物联网资源建模到业务流程设计，再到自然语言描述生成的完整流程，为物联网应用提供了一种高效、智能的建模方法。具体应用场景如光纤授时系统，验证了该方法的有效性和实用性，为未来物联网业务流程建模提供了新的思路和技术支持。

## **资源模型信息结构图**

下面以光纤授时应用场景为例。



## **资源模型**

OWL 作为最常用的本体开发语言之一，能够描述概念类、实例、关系属性、数据属性等信息，具有强大的语义表达能力，在本文中被用作资源模型的语言载体。在构建知识本体时，有几个比较重要的概念：模板类（OWL Class）是对资源类型的抽象；实例（Individual）代表了某类型资源的具体对象；关系属性（Object Property）用于表示主客体事物之间的联系；数据属性（Data Property）用于定义实例的属性值，支持多种数据类型，也能设置值域，或指定适用的实例类别。

下面以光纤授时应用场景为例，介绍物联网资源模型的构建方案与实际应用。我们将资源建模的步骤划分为：搜集领域知识、设计资源模板、创建资源实例、绑定数据属性、设置关系属性等。首先在光纤授时系统中，搜集整理与建模相关的资源信息，完成准备工作。随后，使用 Protégé完成资源模板的设计工作，包括：模板类、关系属性、数据属性的定义等。设计资源模板是为了形式化定义出资源建模的约束，例如：所有设备的类型信息、各类设备可持有的观测属性种类、各种资源间的相互关系，允许绑定在各类实例上的数据属性等。通过对资源进行统一规范的定义描述，制作出可共享、复用的知识本体，进而更高效地完成资源实例信息的构建。

### （1）定义模板类

模板类（OWL Class）是对光纤授时系统中资源类型的抽象。根据本体推荐标准和授时系统资源结构，本文主要将资源分为三类，详细设计如下表所示。



Feature of Interest 根类描述了四类光纤授时设备，在每个分系统中各包含不同型号的设备种类。例如，在“光纤光频传递分系统设备”类别中包含：EDFA中继器、光纤稳频激光器单元、光频传递系统发送设备等抽象类，每类设备均可对应组态软件中的一种设备图元。

Observable Property 根类描述了四类光纤授时设备的观测属性，在每个大类中各包含了多种参量类别。例如，在“光纤时间同步分系统属性”中包含：再生光输出功率、相位噪声、离子泵电流、透射峰电压等抽象类。

Site 根类描述了光纤授时系统中的四类站点，每级站点类各包含了不同类型的子站。例如，在“三级站”中包含：普通站、监测站等抽象类。

### （2）定义关系属性

Object Property（关系属性）涉及主体和客体，用于描述模板类之间或资源实例之间的关系，也可以通过注解来记录更多语义信息。在光纤授时系统中，关系属性的说明如下表所示。



在资源建模时，关系属性的使用方式如下：

（1）为各类设备添加类型表达式，通过“has\_property”关系定义出它可持有的观测属性类别；

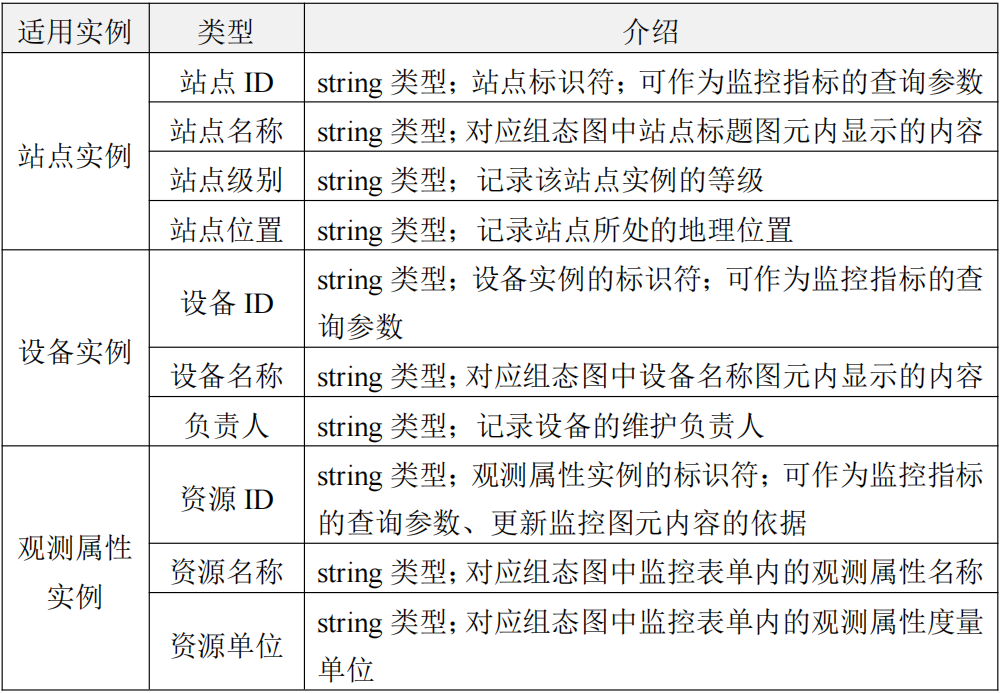
（2）用户创建一个“EDFA 中继器”实例，并将“has\_property”设置在它和一个“设备运行状态”实例上，即代表该设备实例持有一个观测属性实例；

（3）用户为“EDFA中继器”实例和“设备运行状态”实例设置了 trigger\_on 关系属性，并添加“equal”类型的语义注解，内容为“value=0;status=关闭”，那么当组态图处于监控运行态时，若“设备运行状态”的值为 0，则“EDFA 中继器”实例所对应的设备图元展示为“关闭”状态；

（4）用户为“EDFA 中继器”实例与“监控单元”实例设置了 connect 关系属性，并添加“subjectPort”类型的注解，内容为“cellType= 磁力点;connectionName=输出接口”，则代表此关系的主体设备由名为“输出接口”的磁力点与“监控单元”设备实例相连。

### （3）定义数据属性

数据属性用于描述资源实例所绑定的数据信息，可以通过定义域来限制它们的使用范围。它支持多种数据类型，如 string、boolean、int 等。在构建物联网资源模型时，可以为设备实例绑定“设备 ID”和“负责人”等数据属性，或为观测属性实例绑定“资源单位”等，从而更详细地描述观测属性资源。详细设计如下表所示。

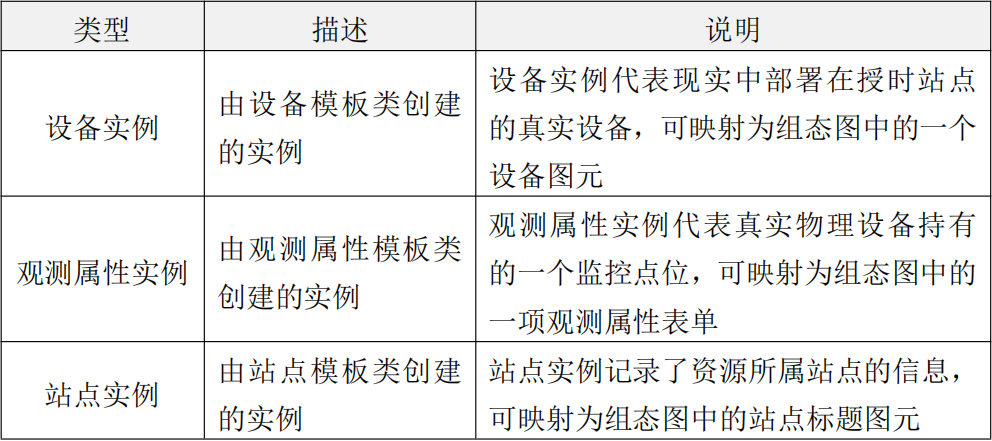


在 Protégé 中定义好上述信息，便完成了资源模板的设计工作，在线建模工具可将资源模板中的内容展示为下图中的结构。



### （4）创建实例

Individual（实例）映射了现实中具体的资源对象，也是 Object Property（关系属性）和 Data Property（数据属性）的主要承载者。根据模板类的定义，物联网资源模型中的实例主要分为三类，详细设计如表所示。



完成上述步骤后，再为资源实例设置对象关系并绑定数据属性，就完成了资源建模工作。构建好的物联网资源模型记录了站点信息、设备实例信息及其持有的观测属性、设备间的光纤连接关系等，可被建模工具展示为下图的结构。

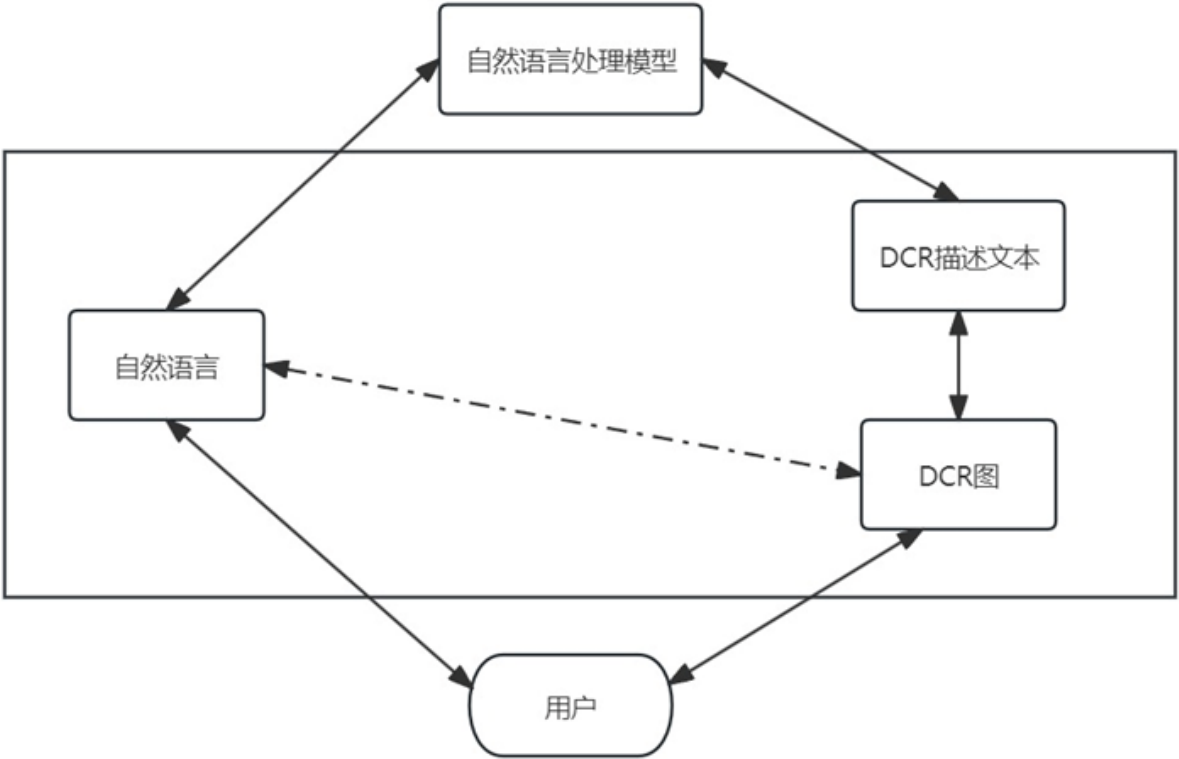


## **业务流程图**

本文研究的流程图与命令式流程图不同，它被分类为声明式流程图，更侧重于**描述做什么而不是怎么做**，事务的执行过程被抽象为一系列的声明或规则，这些声明或规则描述了事务执行的逻辑结构、约束条件和行为特征，而不涉及具体的实现细节。本文的具体研究对象为属于声明式流程图的 **DCR 图**，全称为 Dynamic Condition Response Graphs，直译为中文是动态条件响应图，它是一种用于描述业务流程和工作流程的图形化表示方法，其特点是具有动态性、条件性和响应性。DCR 图能够灵活地表达业务规则和约束条件，并根据特定的条件和事件动态地调整流程的执行路径。由于 DCR 图主要用于描述业务工作流程，因此它相比于一般的声明式流程图而言减少了数据节点这一元素，只留下了声明节点 Node 和连接线 Edge，于是 DCR 图的基本元素如下：

* 节点代表了业务流程中的各个活动或事件。这些节点可以是任务、决策点、状态转换等，用于表示业务流程中的不同阶段和行为。每个节点都具有与之相关联的属性，例如状态（已完成、未完成）、触发条件等。
* 连接线用于连接不同的节点，表示业务流程中的流转路径。连接线上带有标签，用于描述节点之间的约束条件或触发条件。

## **整体设计**



## **模型选择**

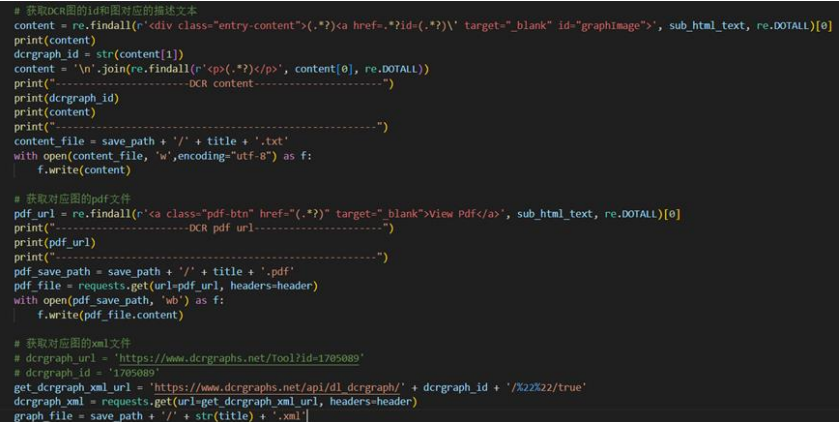
LLAMA（Language Learning Assisted by Machine and Algorithms）模型，它是一种基于深度学习和自然语言处理技术的预训练语言模型。它由百度研究院提出，并基于大规模文本数据进行了预训练。

## **模型构建**

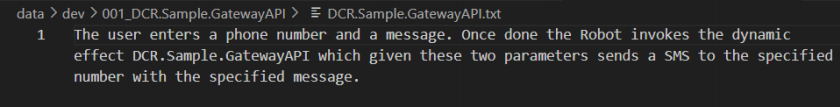
根据选择的 LLAMA 模型结构，在深度学习框架中构建模型。本文选择使用飞桨平台的 PaddlePaddle 深度学习框架来实现模型。PaddlePaddle（Parallel Distributed Deep Learning）是由百度开发的开源深度学习框架，旨在为用户提供高效、灵活、易用的深度学习工具和库。

## **爬虫设计**

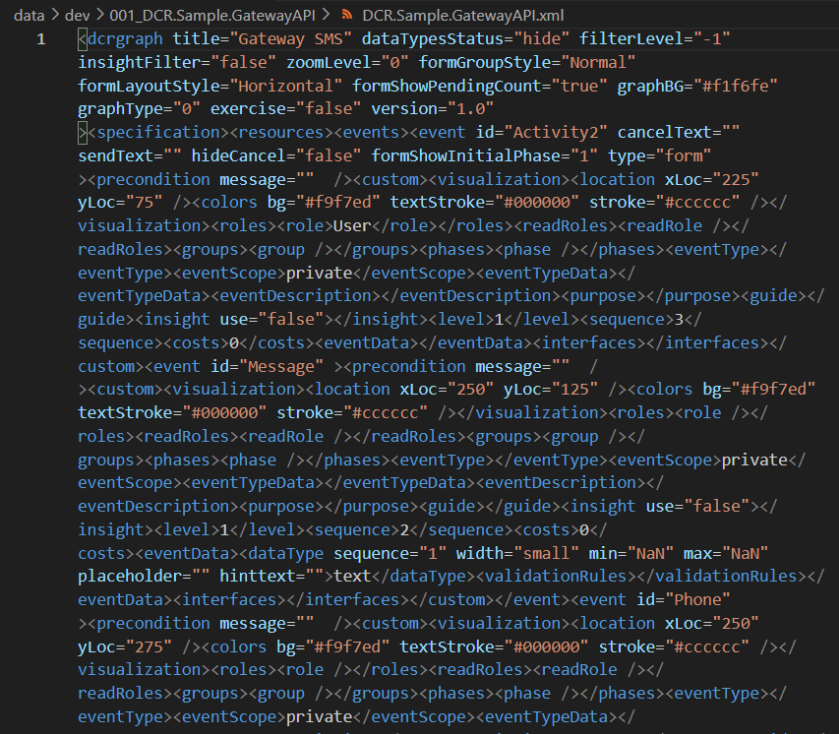
通过设计一个爬虫来爬取相关数据



该爬虫每爬取一个完整 DCR 图用例就会根据网页源码生成两个文件保存在本地，一个是该 DCR 图的自然语言描述的 TXT 文件：



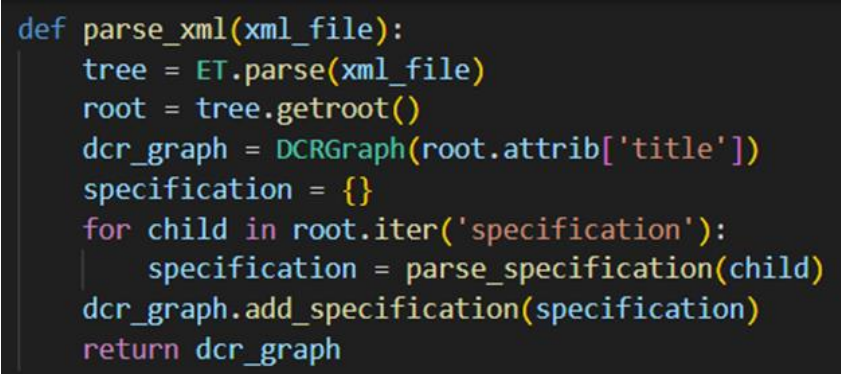
另一个则是该描述 DCR 图的 XML 文件：

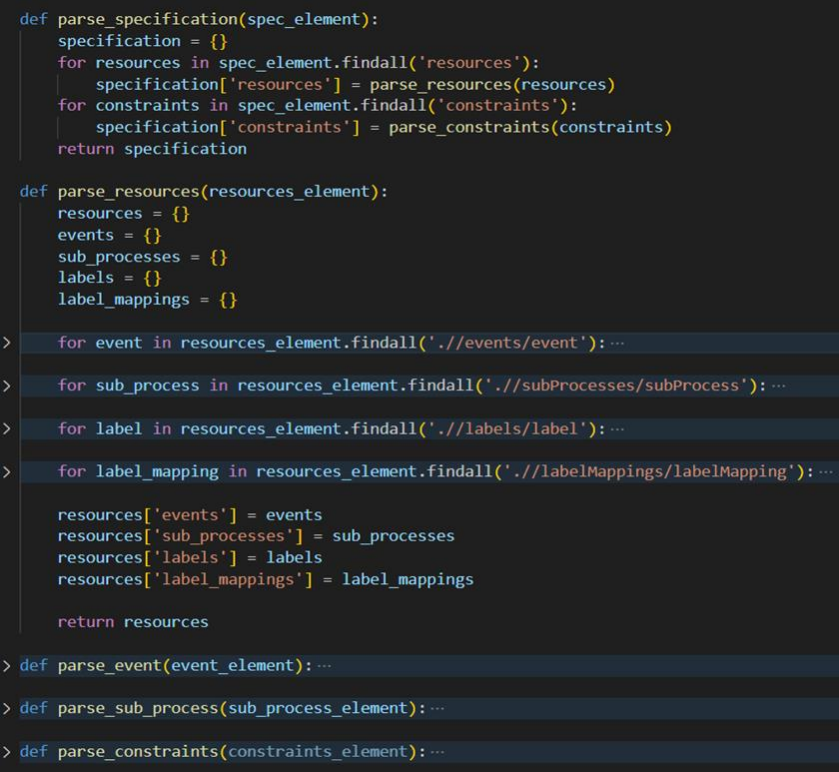


为了方便模型的训练，需要将爬虫爬取的 XML 文件中零散的信息提取出来，并转换为设计好的标准数据结构，这样就有了 DCR 图描述文本格式的 JSON 文件：

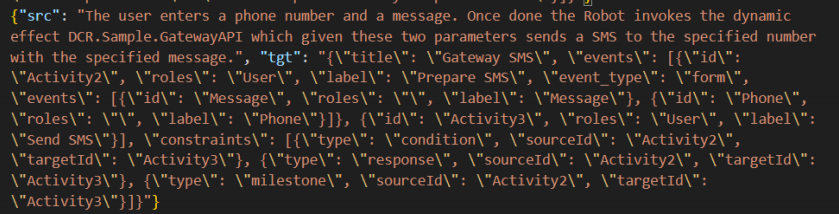


XML -> JSON 部分代码截图如下：

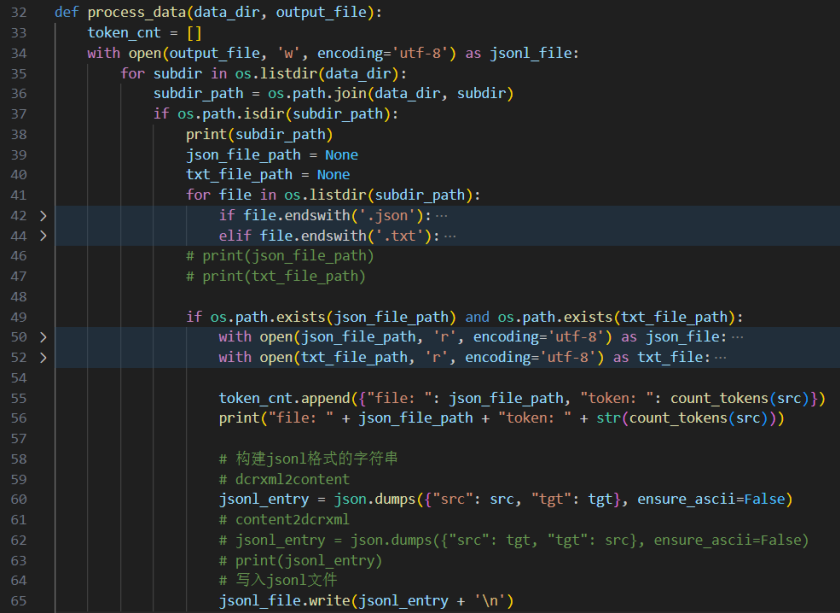




再将 TXT 文件中的自然语言描述文本和 DCR 图描述文本放在同一个 JSON 文件中，这样就制作好了一条数据集中的标准数据：



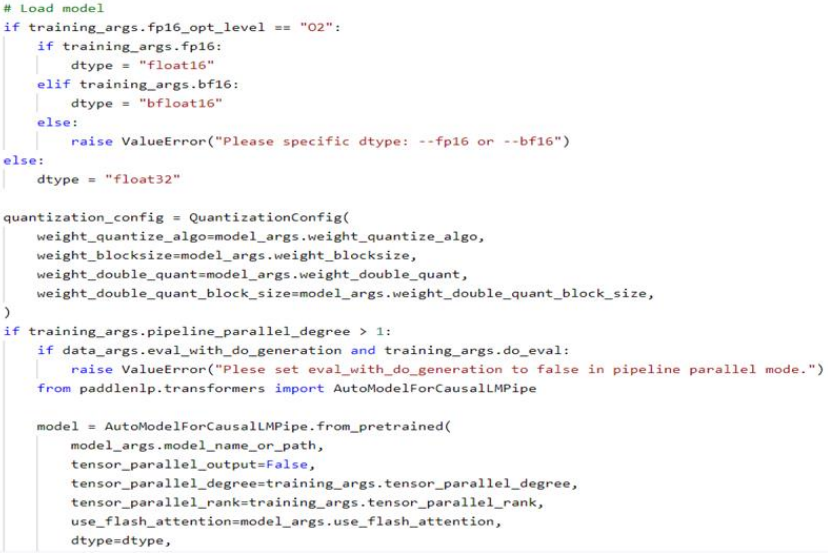
JSON + TXT -> JSON 部分代码如下：

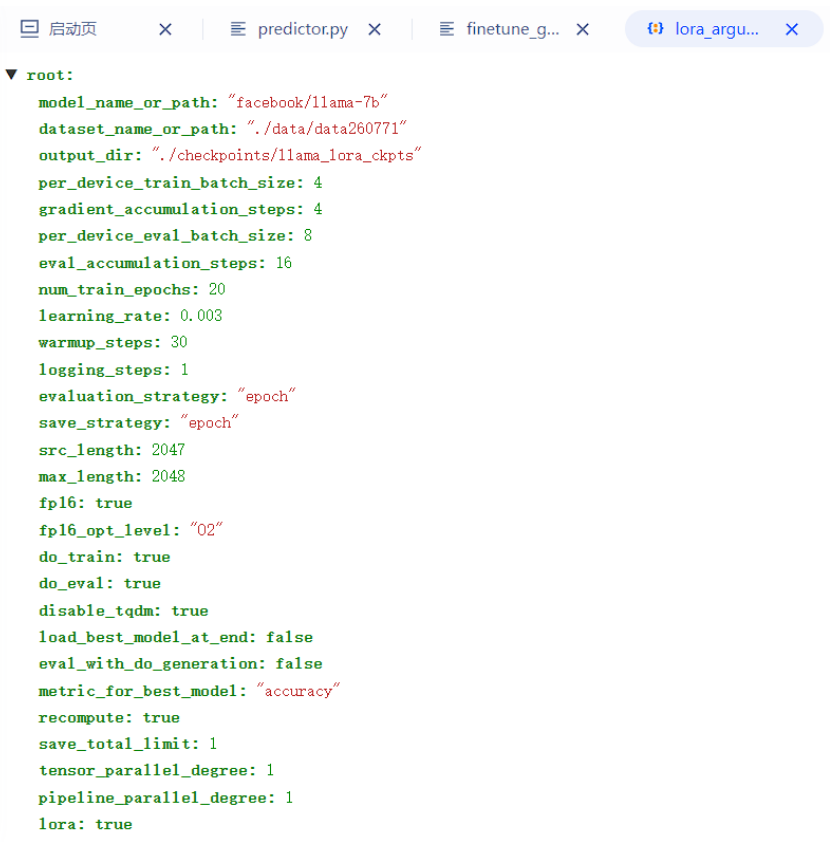


综上所述，先将所有的 XML 文件转换为包含零散 DCR 图信息的 JSON 文件，然后提取 JSON 文件信息按照标准 DCR 图描述文本格式与自然语言文本一同按照训练数据集格式存在一个新的 JSON 文件中，最后就得到了一个充满了训练数据集格式的 JSON 文件夹。

## **模型训练细节**

该部分主要介绍各种参数在训练模型中的作用





1、model\_name\_or\_path: "facebook/llama-7b"

指定要使用的预训练模型的名称或路径。在这里，模型使用的是 Facebook 发布的LLAMA 模型的 7B 版本，作为预训练模型。

2、dataset\_name\_or\_path: "./data/data260771"

指定要加载的训练数据集的名称或路径。这里的数据集路径是"./data/data260771"，这是我上传的包含自然语言文本和对应的 DCR 描述文本的数据集。

3、output\_dir: "./checkpoints/llama\_lora\_ckpts"

指定模型训练过程中保存检查点和输出文件的目录路径。在这里，模型的检查点和输出文件将保存在"./checkpoints/llama\_lora\_ckpts"目录中。

4、per\_device\_train\_batch\_size: 4

指定每个训练设备的批量大小。在训练过程中，每个设备处理的样本数量为 4。

5、gradient\_accumulation\_steps: 4

指定梯度累积的步数。在每个训练步骤中，梯度更新将累积 4 个步骤的梯度后再进行更新。

6、per\_device\_eval\_batch\_size: 8

指定每个评估设备的批量大小。在评估过程中，每个设备处理的样本数量为 8。

7、eval\_accumulation\_steps: 16

指定评估累积的步数。在每个评估步骤中，结果将累积 16 个步骤的结果后再进行评估。

8、num\_train\_epochs: 20

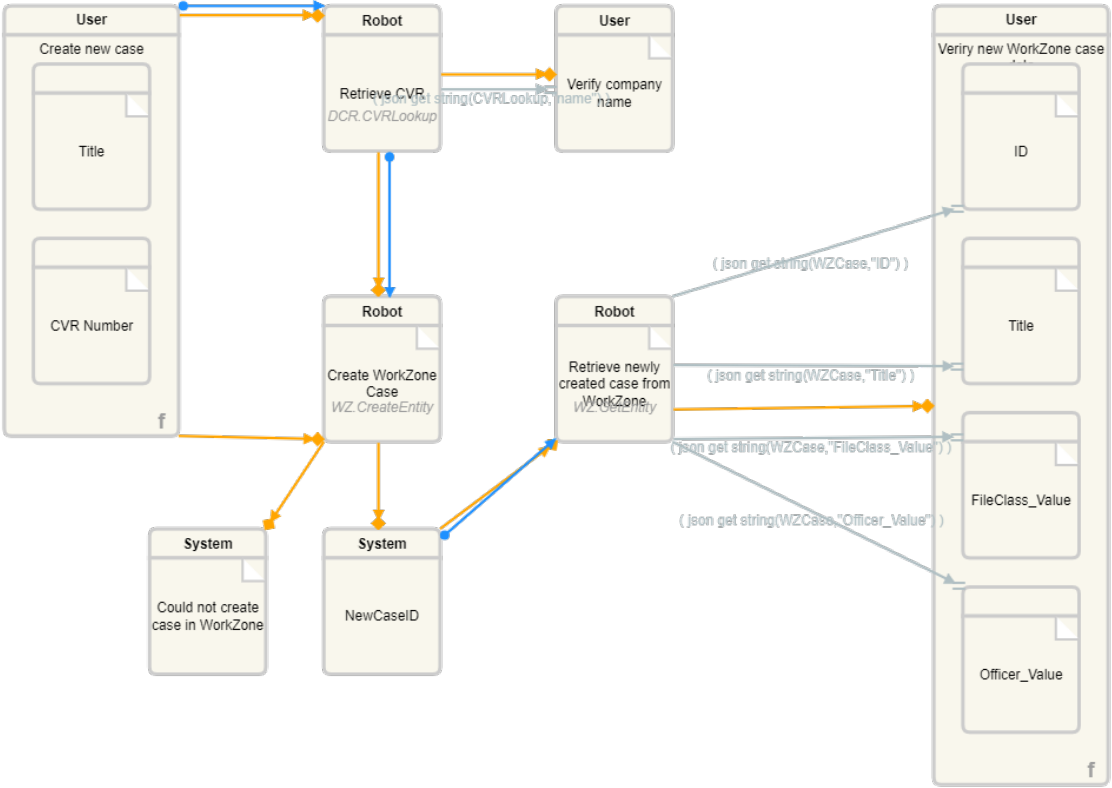
指定训练的总轮数。模型将被训练 20 个 epoch。

9、learning\_rate: 0.003

指定学习率的初始值。学习率初始化为 0.003。

## **例子**

### （1）DCR图



### （2）自然语言描述

First the user must fill out the form values Title and CVR Number. The the robot lookup the CVR Number using the DCR.CVRLookup effect. The user can inspect the company once completed.  
Using the company name a new case is created in WorkZone using the WZ.CreateEntity effect. In case it succeeds the id of the WorkZone case is returned, and information about the newly created case is retrieved using WZ.GetEntity. Once completed the user can inspect the values.  
If an error happens or the request times out, the activity Could not create case in WorkZone is executed indicating that an error happened.

### （3）DCR描述文本（json）

{  
 "title": "WZ.CreateEntity - Create WorkZone case with part being a CVR number - CVRLookup, CreateEntity",  
 "events": [  
 {  
 "id": "Form",  
 "roles": "User",  
 "label": "Create new case",  
 "event\_type": "form",  
 "events": [  
 {  
 "id": "CaseTitle",  
 "roles": "",  
 "label": "Title"  
 },  
 {  
 "id": "CVRNumber",  
 "roles": "",  
 "label": "CVR Number"  
 }  
 ]  
 },  
 {  
 "id": "CVRLookup",  
 "roles": "Robot",  
 "label": "Retrieve CVR"  
 },  
 {  
 "id": "CreateCase",  
 "roles": "Robot",  
 "label": "Create WorkZone Case"  
 },  
 {  
 "id": "CompanyName",  
 "roles": "User",  
 "label": "Verify company name"  
 },  
 {  
 "id": "NewCaseID",  
 "roles": "System",  
 "label": "NewCaseID"  
 },  
 {  
 "id": "WZCase",  
 "roles": "Robot",  
 "label": "Retrieve newly created case from WorkZone"  
 },  
 {  
 "id": "Activity9",  
 "roles": "User",  
 "label": "Veriry new WorkZone case data",  
 "event\_type": "form",  
 "events": [  
 {  
 "id": "Activity12",  
 "roles": "",  
 "label": "Officer\_Value"  
 },  
 {  
 "id": "Activity11",  
 "roles": "",  
 "label": "FileClass\_Value"  
 },  
 {  
 "id": "Activity10",  
 "roles": "",  
 "label": "Title"  
 },  
 {  
 "id": "Activity13",  
 "roles": "",  
 "label": "ID"  
 }  
 ]  
 },  
 {  
 "id": "CreateCaseFailed",  
 "roles": "System",  
 "label": "Could not create case in WorkZone"  
 }  
 ],  
 "expressions": [  
 {  
 "id": "CVRLookup-path-CompanyName--update--value",  
 "value": "json get string(CVRLookup,"name")"  
 },  
 {  
 "id": "WZCase-path-Activity10--update--value",  
 "value": "json get string(WZCase,"Title")"  
 },  
 {  
 "id": "WZCase-path-Activity11--update--value",  
 "value": "json get string(WZCase,"FileClass\_Value")"  
 },  
 {  
 "id": "WZCase-path-Activity12--update--value",  
 "value": "json get string(WZCase,"Officer\_Value")"  
 },  
 {  
 "id": "WZCase-path-Activity13--update--value",  
 "value": "json get string(WZCase,"ID")"  
 }  
 ],  
 "constraints": [  
 {  
 "type": "condition",  
 "sourceId": "Form",  
 "targetId": "CVRLookup"  
 },  
 {  
 "type": "condition",  
 "sourceId": "Form",  
 "targetId": "CreateCase"  
 },  
 {  
 "type": "condition",  
 "sourceId": "CVRLookup",  
 "targetId": "CreateCase"  
 },  
 {  
 "type": "condition",  
 "sourceId": "WZCase",  
 "targetId": "Activity9"  
 },  
 {  
 "type": "condition",  
 "sourceId": "CreateCase",  
 "targetId": "NewCaseID"  
 },  
 {  
 "type": "condition",  
 "sourceId": "NewCaseID",  
 "targetId": "WZCase"  
 },  
 {  
 "type": "condition",  
 "sourceId": "CVRLookup",  
 "targetId": "CompanyName"  
 },  
 {  
 "type": "condition",  
 "sourceId": "CreateCase",  
 "targetId": "CreateCaseFailed"  
 },  
 {  
 "type": "response",  
 "sourceId": "Form",  
 "targetId": "CVRLookup"  
 },  
 {  
 "type": "response",  
 "sourceId": "CVRLookup",  
 "targetId": "CreateCase"  
 },  
 {  
 "type": "response",  
 "sourceId": "NewCaseID",  
 "targetId": "WZCase"  
 },  
 {  
 "type": "milestone",  
 "sourceId": "Form",  
 "targetId": "CVRLookup"  
 },  
 {  
 "type": "milestone",  
 "sourceId": "Form",  
 "targetId": "CreateCase"  
 },  
 {  
 "type": "milestone",  
 "sourceId": "CVRLookup",  
 "targetId": "CreateCase"  
 },  
 {  
 "type": "milestone",  
 "sourceId": "WZCase",  
 "targetId": "Activity9"  
 },  
 {  
 "type": "milestone",  
 "sourceId": "CreateCase",  
 "targetId": "NewCaseID"  
 },  
 {  
 "type": "milestone",  
 "sourceId": "NewCaseID",  
 "targetId": "WZCase"  
 },  
 {  
 "type": "milestone",  
 "sourceId": "CVRLookup",  
 "targetId": "CompanyName"  
 },  
 {  
 "type": "milestone",  
 "sourceId": "CreateCase",  
 "targetId": "CreateCaseFailed"  
 },  
 {  
 "type": "update",  
 "sourceId": "CVRLookup",  
 "targetId": "CompanyName",  
 "valueExpressionId": "CVRLookup-path-CompanyName--update--value"  
 },  
 {  
 "type": "update",  
 "sourceId": "WZCase",  
 "targetId": "Activity10",  
 "valueExpressionId": "WZCase-path-Activity10--update--value"  
 },  
 {  
 "type": "update",  
 "sourceId": "WZCase",  
 "targetId": "Activity11",  
 "valueExpressionId": "WZCase-path-Activity11--update--value"  
 },  
 {  
 "type": "update",  
 "sourceId": "WZCase",  
 "targetId": "Activity12",  
 "valueExpressionId": "WZCase-path-Activity12--update--value"  
 },  
 {  
 "type": "update",  
 "sourceId": "WZCase",  
 "targetId": "Activity13",  
 "valueExpressionId": "WZCase-path-Activity13--update--value"  
 }  
 ]  
}

### （4）推理结果示例

