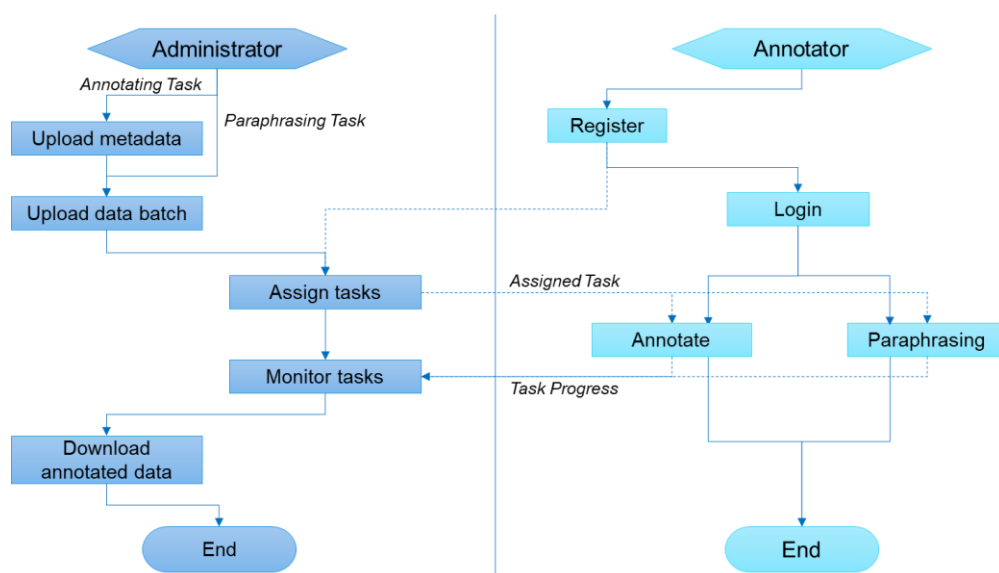# metaCAT Administrator Operation Guide

metaCAT is an open-source web-based annotation tool designed specifically for developing task-oriented dialogue data. metaCAT is an annotation tool that provides comprehensive metadata annotation coverage to the domain, intent, and span information. The data annotation quality is enhanced by a real-time annotation constraint-checking mechanism. The annotation process is further streamlined to create more diversified and robust annotated utterances by providing an ASR-inclusive paraphrasing function.

1   Working Flow

The working flow of metaCAT is as following:

The flow elements related to the administrator are as below：

Annotating Task: To annotate dialogue intents and slots based on the dialogue data which has been manually collected through crowd-sourcing.

Paraphrasing Task: To paraphrase dialogue utterances based on the standard dialogues which are automatically generated using rule-based algorithm via keyboard typing or ASR inputting. Both

utterances from "Usr" and "Sys" sides should be paraphrased using spoken natural language. And then some slots should be re-annotated according to their new span information inside paraphrased utterances.

Upload Metadata: To upload metadata, which describe the ontology of annotating data. Metadata can be applied only on annotating task. Please refer to related provided sample file to get detail file format.

Upload Data Batch: To upload compressed packed data file, which includes batches to be assigned to annotators. Please refer to related provided sample file to get detail file format.

Assign Tasks: To assign tasks (i.e. batches) to registered annotators after the system has extracted batches from uploaded data.

Monitor Tasks: To monitor the performing schedule and status of the assigned tasks.

Download Data Batch: To download the result data of annotating or paraphrasing task (partly finished or totally finished).

2 Module Introduction

2.1 System Operation

This module provides functions of initialization, importing and exporting of data.

Directory Initialization(①): To initialize the file storage of the system and MongoDB database.

Metadata Importing(②): To import metadata for annotating tasks. Please refer to related provided sample file to get detail file format.

Batch Importing(③④): To upload data file for annotating or paraphrasing tasks. Two types of file format are supported: JSON batch and RAW batch (for annotating task only). Please refer to related provided sample file to get detail file formats.

Batch Exporting(⑤⑥⑦): To download result data file according to the assigned and being performed annotating or paraphrasing tasks. Three types of format are supported: annotating batch, paraphrasing batch and MultiWOZ batch (Auto-transferred to MultiWOZ format). The three export formats are consistent with the import format, and related task progress information is also attached inside.

Operation Error(⑧): If any error occurs during the above operations, the system will show all error records so that for the administrator can check and fix them.

2.2 Data Management

This module provides functions of data displaying and task assignation. It is often used to collaborate with the system operation module.

Dataset Metadata (①)**:** List of metadata corresponding to the dataset.

View Metadata(②)**:** To click to view the JSON data format of metadata.

Metadata Overview(③)**:** Describes the metadata used by metaCAT and the dependency between metadata elements.

Annotator Selection(④)**:** To select a registered annotator for task assignation.

Task Selection(⑤)**:** To select an annotating task or paraphrasing task to assign to the selected annotator.

Batch Assignation(⑥)**:** To assign the selected tasks (i.e. batches) to the selected annotator.

No.	Metadata Category	Metadata Name	Operation
1	annotating ①	MULTIWOZ	View ②
2	annotating	VehicleSD	View

Metadata is a segment of data used to describe the dialogue domain, intention, and slot annotation system. It is stored in JSON format and determines the optional service domains on the user side and system side, slots in each service domain, valid slot-values in each slot, and allowed intents.

★ **General:** This type of domain contains only some intent without actual slots, such as "Thank" and "Bye", etc.

★ **Service Domain:** These types of domain contain some specific intents with slots, e.g. "Hotel" and "Restaurant". Generally, a dialogue involves one service domain. However, in a few cases, it may involve two service domains, e.g. "Hotel" and "Taxi".

★ **Intent:** It represents the intent expressed in each utterance. It is possible that one utterance contains multiple intents.

★ **Slot:** It indicates the key information carried by the intent. Non-enumerated slots are usually one word or short segment of text taken from the original utterance text. Enumerated slots only have some specific slot-values and may not appear in the utterance text of the dialogue.

2.3 Annotating Progress

This module displays the performing schedule and status of assigned annotating tasks or paraphrasing tasks.

No.	User Name	Batch No.	Metadata Category	Metadata Name	Task Progress	Description
1	Bill	json_annotating002	annotating	MULTIWOZ	20%	Initialized
2	Bill	raw_annotating001	annotating	MULTIWOZ	0%	Initialized
3	Bill	paraphrasing001	paraphrasing	NA	0%	Initialized

2.4 System Setting

This module is used to change the administrator password and set the switch of each function.

Administrator Password Changing(①): To change the password of the administrator. Please refer to related description on GitHub of metaCAT to get the default password.

Switch for System Annotating(②): Whether to support "System" side annotating. Opened by default.

Switch for Dialogue Turn Deletion(③): Whether to allow dialogue turn deletion. Closed by default.

Switch for Dialogue Utterance Creation (④): Whether to allow dialogue utterance creation (appended to end of the dialogue). Closed by default.

Switch for ASR Input (⑤): Whether to allow ASR (Automatic Speech Recognition) input. Closed by default.

3   Revise History

Version No.	Revision Date	Revision Summary
1.0	2020/07/31	First draft.