FALL 2020 – CSE3063 Object Software Design

Java Project Iteration #1

Group 10

Description

Data labeling is the process of detecting and tagging data samples to be used in machine learning. In this project we analyze, design and implement a data labeling system in an object oriented manner.

In this iteration we created an object based design that includes reading and writing datasets, taking users and a random labeling mechanism.

Glossary

Term	Definition
Dataset	The collection of labels and instances
Label	An informative tag that can be used in assignments
Instance	A data sample to label
User	Someone who uses labeling mechanism
Labeling Mechanism	Process of assigning labels to instances
Assignment	The allocation of a label or multiple labels to a group of instances.
Random Labeling Mechanism	A label mechanism that choses a random label and assigns to an instance.

Requirements

- Reading dataset files with labels and instances

- Reading configuration files to add users

- Supporting multiple users

- A random labeling mechanism

- Printing actions to the command line and a log file

- Each user will label all the instances one by one

DatasetController

operates

Dataset
- id
- name
- instance type
- labels
- instances
- maximum number of labels per instance

Instance
- id
- label ids
- name

contains

Label
- id
- text

gets stored

contains

Random Labeling Mechanism
- name

User
- id
- name
- user type

uses

sub mechanism

Mechanism
- name

uses

creates

Assignment
- instance id
- user id
- date
- labels

gets stored

gets stored

labels

uses

Log

Storage
- dataset
- users
- assignments