JAVA LABELING SYSTEM

Software Requirement Analysis
V1.0

Lead Software Developers

Mahmut Hilmi ARIKMERT
Muhammed Enes AKTÜRK
Yunus Emre ERTUNÇ
Rabiul ISLAM
Asaf Talha GÜLTEKİN
Hamza TÜRKMEN
Kerim BOYACI

Customers

Murat Can GANİZ
Lokman ALTIN

Introduction

The purpose of this project is creating a data labeling system for different NLP problems using a console based program. Industry standard Object Oriented Designing and Programming will be implemented for this project. The chosen language is Java and agile software development methodology will be followed throughout the project.

Overview

Data labeling means assigning the several predetermined labels (class labels, categories, tags) to a group of instances (samples, examples, records, documents). Firstly, for implementing this system we will use a dataset in json file format. After necessary operations we will produce an output file. The dataset will contain set of labels, instances, labeling properties of instances and user information. We will use random labeling mechanism that will select a label randomly and then assign it to an instance. An instance can be assigned more than one label. After this process the output file will be produced that will include matchups.

Functional Requirements

- **Usability Requirements**

  a. The app will be a console based program
  b. The data set will be pointed through console
  c. The data set should be readable and easily accessible
  d. Multiple users will be assigned randomly
  e. The user information will be predefined in json file
  f. The default problem type will be assigned to user randomly
  g. Labeling is randomly chosen.
  h. The action should be logged and printed on the command line
  i. The output of the result will be recorded in a json file.
  j. The result should be as accurate as possible
  k. The app will be work in English language
  l. Type of error handling will be error messages

- **Implementation Requirements**

  a. The app will be written in Java with OOP concept.
  b. The app will take json file as input
  c. There will be no database integrity for this project

- **Physical Requirements**
  a. The app can run without login
  b. The app doesn’t need to create any user
  c. The app doesn’t work on web browser
  d. The app should be light and easy to operate through console

Non-Functional Requirements

- **Performance Requirements**

  a. The app size should be medium neither too large nor too small
  b. The app should response quickly
  c. If any problem or exception detected, it should be handled in a short time, and recover the system.
  d. Resource usage should not be too much.

- **Supportability Requirements**

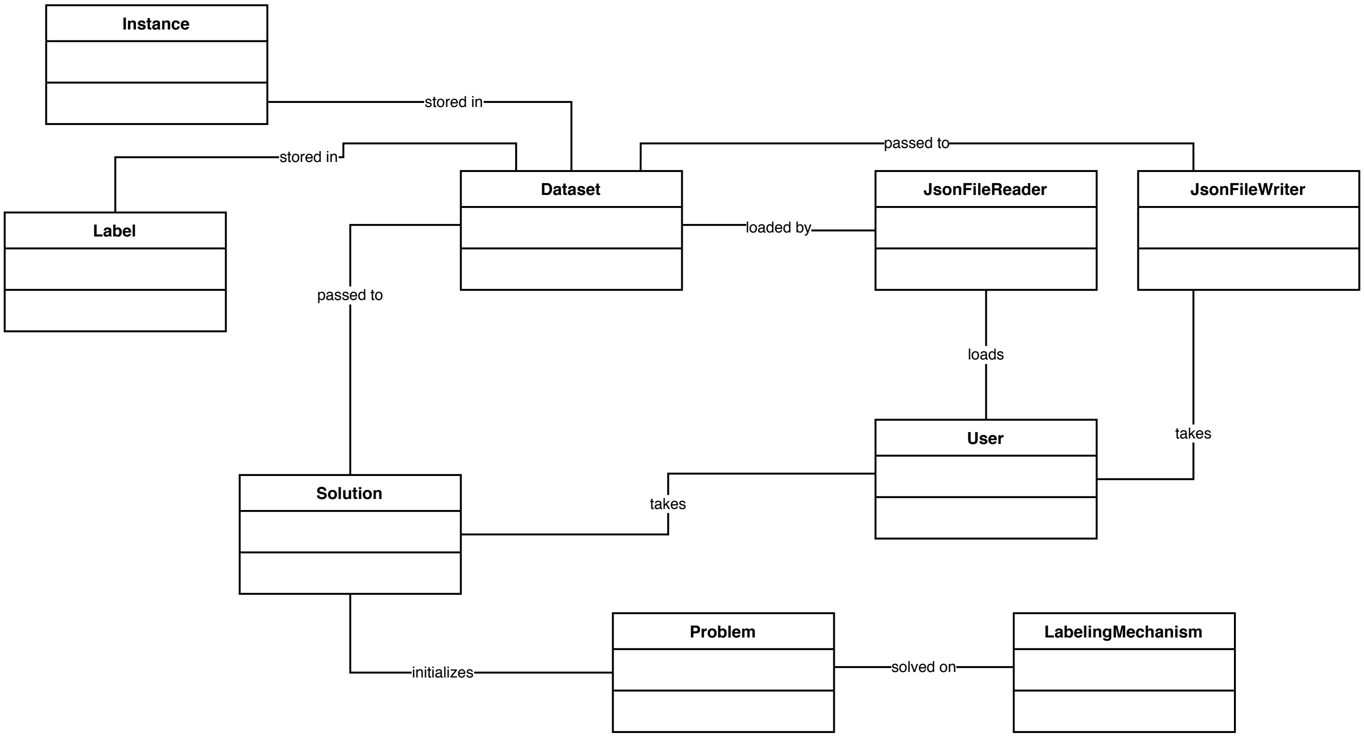
  a. The app should be extensible and changeable according to client’s expectations.
  b. The app should be adaptive for different kind of situations.
  c. The app should have compatibility for different OS.

- **Reliability Requirements**

  a. The app should perform as expected from command line

b. The app should produce result as accurate as possible
c. The app shouldn't lag or throw error halfway of the operation.

Domain Model



Glossary

NLP – Natural Language Processing
Sentiment Analysis – Understanding human language's emotion
Label – Categorizing human language
Console – Command Line or a black window that is seen in windows/linux operating system. User needs to type through keyboard to provide necessary instruction to the program instead of normal mouse click and navigation button.
Json – JSON stands for JavaScript Object Notation. JSON is a lightweight format for storing and transporting data