

Code Problem Similarity Checker

An automated tool to detect similar algorithmic problems from source codes

Student: Yeo Yun Siang Geremie Supervisor: Anwitaman Datta

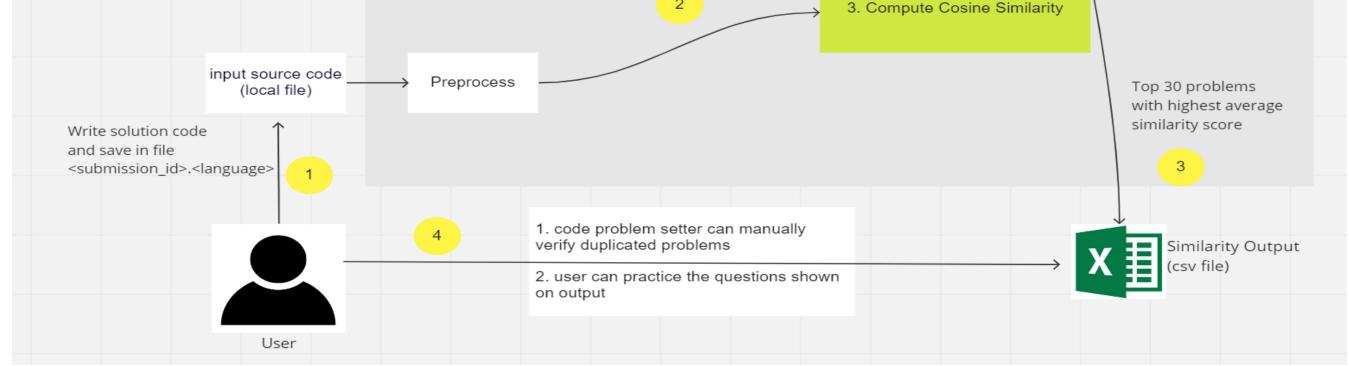
Project Objectives:

Created a new dataset of 107k source codes in 3 different languages from about ~3k different problems in Codeforces.

Experimented code clone detection using 3 different pre-trained models (C4-CodeBERT, GraphCodeBERT, UniXcoder)

Designed the pipeline for the similarity checker utilizing the best pre-trained model from the experiments

Code Problem Similarity Checker
Codeforces dataset (107k source codes) Preprocess
UniXcoder (modified) model 1. Code Embeddings (input) 2. Code Embeddings (dataset)



Code Problem

F. Souvenirs
time limit per test: 3 seconds
memory limit per test: 512 megabytes
input: standard input
output: standard output

Artsem is on vacation and wants to buy souvenirs for his two teammates. There are n souvenir shops along the street. In *i*-th shop Artsem can buy one souvenir for a_i dollars, and he cannot buy more than one souvenir in one shop. He doesn't want to introduce envy in his team, so he wants to buy two souvenirs with least possible difference in price.

Artsem has visited the shopping street *m* times. For some strange reason on the *i*-th day only shops with numbers from l_i to r_i were operating (weird? yes it is, but have you ever tried to come up with a reasonable legend for a range query problem?). For each visit, Artsem wants to know the minimum possible difference in prices of two different souvenirs he can buy in the opened shops.

In other words, for each Artsem's visit you should find the minimum possible value of $|a_s - a_t|$ where $l_i \le s, t \le r_i, s \ne t$.

Input

The first line contains an integer $n (2 \le n \le 10^5)$.

The second line contains *n* space-separated integers $a_1, ..., a_n$ ($0 \le a_i \le 10^9$).

The third line contains the number of queries $m (1 \le m \le 3 \cdot 10^5)$.

Next *m* lines describe the queries. *i*-th of these lines contains two space-separated integers l_i and r_i denoting the range of shops working on *i*-th day ($1 \le l_i \le r_i \le n$).

Output Print the answer to each query in a separate line.

Source Code 1

Source Code 2



Problem 1

Problem Similarity Checker

{Problem 2, Problem 3, ...}

https://www.ntu.edu.sg/scse