MASTER THESIS

Datamining on an online judge

Author:
Maxime MARLIER

Supervisor:
Jordi PETIT

A thesis submitted in fulfillment of the requirements
for the degree of

in the

April 26, 2016

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Abstract

Facultat d'Informatica de Barcelona
Master in Innovation and Research in Informatics

Datamining on an online judge

by Maxime MARLIER

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Contents

List of Figures

List of Tables

Chapter 1

Introduction, motivation and goals

Chapter 2

State of the art

Chapter 3

Methodology

3.1 Database description

Here is the list of the table in the database:

3.1.1 users

```
         users
◆user_id        text
•creation_date  date
•administrator  int
•instructor     int
•demo           int
°unregistered   int
```

Description:
The first table contains the users. For this analysis, the user table has been anonymized. We only refer to a user ID, and his contributions in the data base. Personal data from users will not be used for analysis. Only the creation date is kept for a time based analysis.

However, it's needed to exclude some *non-representative* users :

- Some users used for developement ([list])
- Users with a id patern different that *Uxxxxx* (Users used for competition for exemple)
- Demonstrations users ($demo == 1$).
- Instructors, administrators and unregistred users (cf flags atributes in the database).

Numbers:
In term of numbers, the database contains:

- 10565 users in total.
- 55 unregistred users.
- 50 instructors.
- 7 administrators.
- 1 demo user.

3.1.2   problems and abstractproblems

problems	
◆**problem_id**	**text**
•problem_nm	text
•language_id	text
•title	text
•original_language_id	text
•checked	int

Description:
A single problem could be proposed in various languages but the language variation doesn't affect the technical details of a same problem. That means that the way how a submission would be processed is never linked to the language[1].

That explains those two tables discribing the problems. The first one colled *abstractproblems* contains the technical informations for submission management. The second one, *problems*, is the description of a problem, according to a specific language (*language_ id*) and refering to a *abstractproblems*.

There is different types of problems :

- *Pxxxxx*
  Those type of problems will be our baseline for the analysis. In fact, they are the *offical* problems initialy present in the database, created by the designers. We can consider them as *right* and *relevant* in term of submission and veredict[2].

abstractproblems	
◆**problem_nm**	**text**
•user_id	text
•public	int
•official	int
∘compilers	text
∘deprecation	text
•checked	int

- *Xxxxxx*
  The letter *X* means *externe*. Those problems have been created by users (*instructors*) and havn't been validatd by anyone. Moreover, only a portion of users can acces to it (Those who suscribed to the courses related to the same instructor)

- *Gxxxxx*
  The letter *G* means *game*. Those problems are used on a very specific scenario. There is only a very few of them and they will be ignored in our analysis.

- *deprecated*
  Obviously, this type of problem is not relevant for the analysis.

Numbers:

- 1909 abstractproblems in total.

- 1325 Pxxxxx like abstractproblems.

- 575 Xxxxxx like abstractproblems.

- 9 Gxxxxx like abstractproblems.

- 85 deprecated abstractproblems (including 21 Pxxxx type).

---

[1]There are actually few problems which differ between languages for inputs or outputs regarding to the language but those are negligible

[2]The concept of veredict will be explain in the following section *submissions table*

3.1.3 submissions

submissions	
submission_uid	**text**
•user_id	text
•problem_id	text
•submission_id	text
•compiler_id	text
•state	text
•time_in	timestamp(0)
∘time_out	timestamp(0)
∘veredict	text
∘veredict_info	text
∘internal_error	text
∘legacy	int
∘veredict_publics	text
∘ok_publics_but_wrong	int
∘score	text

Description:
Every instance in this table represents the submission of a solution for a specific problem (*problem_id*) by a specific user (*user_id*) at a given time/moment (*time_in* (*time_out*)). Form that submission (after a internal process) will stand out a veredict meaningful of the submission correctness.

Numbers:

-
-
-

Veredict distribution :

Acronym	Veredict	%
AC	Accepted	43.62
WA	Wrong Answer	30.06
EE	Execution Error	11.41
CE	Compilation Error	10.70
PE	Presentation Error	3.62
SC	Scored	0.30
IC	Invalid Character	0.29
SE	Setter Error	0.01
FE	Fatal Errors	0.00
NC	Noncompliant Solution	0.00
Pending	Pending Submission	0.00
IE	Internal Error	0.00

3.1.4 courseslists

courseslists	
course_id	**text**
list_id	**text**
•position	int

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.
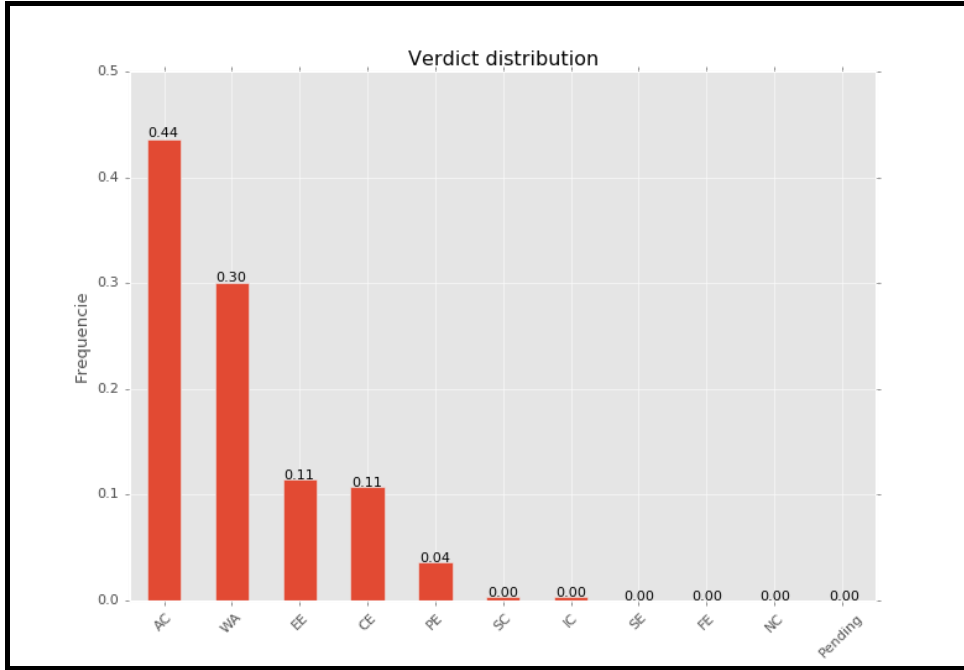
FIGURE 3.1: Frequency distribustion of veredict accros every relevant submissions

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

3.1.5   courses

courses	
course_id	**text**
user_id	text
course_nm	text
title	text
description	text
annotation	text
public	int
official	int
sort_time	timestamp

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

3.1.6   coursesusers

coursesusers	
course_id	**text**
user_id	text
tutor	int
tag	text

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

3.1.7   listitems

```
┌─────────────────────────┐
│        listitems        │
├─────────────────────────┤
│ •list_id          text  │
│ •position          int  │
│ ºproblem_nm       text  │
│ ºdescription      text  │
└─────────────────────────┘
```

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

3.1.8   lists

```
┌──────────────────────────────┐
│             lists            │
├──────────────────────────────┤
│ •list_id          text       │
│ •user_id          text       │
│ •lit_nm           text       │
│ •title            text       │
│ ºdescription      text       │
│ ºannotation       text       │
│ •public           int        │
│ •official         int        │
│ ºshort time       timestamp  │
└──────────────────────────────┘
```

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

3.1.9   problemstags

```
┌─────────────────────────┐
│       problemstags      │
├─────────────────────────┤
│ •problem_nm  text       │
│ •tag         text       │
└─────────────────────────┘
```

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et,

tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Chapter 4

Development of the proposal/technical/work

Chapter 5

Evaluation of the proposal/technical/work

Chapter 6

Conclusions

Appendix A

Appendix

Bibliography

[1] Leslie Lamport, *LATEX: a document preparation system*, Addison Wesley, Massachusetts, 2nd edition, 1994.