Platform for Analysing and Encouraging Student Activity on Contest and E-learning Systems

Bojan KOSTADINOV, Mile JOVANOV, Emil STANKOV
Faculty of Computer Science and Engineering, University Ss. Cyril and Methodius
st. Rugjer Boshkovikj 16 Skopje, Macedonia
e-mail: {bojan.kostadinov, mile.jovanov, emil.stankov}@gmail.com

Abstract. Data collection and machine learning are changing the world. Whether it is medicine, sports or education, companies and institutions are investing a lot of time and money in systems that gather, process and analyse data. Likewise, to improve competitiveness, a lot of countries are making changes to their educational policy by supporting STEM disciplines. Therefore, it’s important to put effort into using various data sources to help students succeed in STEM. In this paper, we present a platform that can analyse student’s activity on various contest and e-learning systems, combine and process the data, and then present it in various ways that are easy to understand. This in turn enables teachers and organizers to recognize talented and hard-working students, identify issues, and/or motivate students to practice and work on areas where they’re weaker.

Keywords: students, platform, STEM, data, e-learning, programming contests.

1. Introduction

Science, Technology, Engineering and Mathematics (STEM) are academic disciplines that have a huge potential to improve competitiveness and spur economic development. Nowadays, many countries are working very hard to change their current education policy, in order to respond to the increased demand for occupations related to science. Even more, the topics related to STEM are important for other jobs and aspects of life – as creative thinking, proofs and observations can be successfully utilized in various fields.

A number of researchers have been working on developing innovative ways to motivate students and help them succeed in STEM (e.g. Hall et al., 2011; Hossain and Robinson, 2012; Wang and Degol, 2013; Joyce, 2014; Kearney, 2016). With the increase in popularity of various open-source software and code-repository hosting solutions, schools and teachers have a relatively easy way of adopting software in education. Likewise, there are open websites and systems that offer tutorials, documentation and tasks that anyone can freely use. In this paper, by mainly focusing on informatics and algo-
rithmic thinking, we’ll discuss the usage of libraries and tools for gathering data and statistics from various e-learning platforms and websites, as well as a way to process that data to produce meaningful results. For example, by analysing student activity, we can identify issues such as lack of motivation or inadequate use of time – thus enabling us to help the affected students succeed better.

The paper is organized as follows. In Section 2 we will discuss education, STEM and competitions in informatics. In Section 3 we focus on various contest and e-learning systems for teaching informatics and preparing students for competitions. We will also describe several popular websites that organize algorithmic contests. Section 4 contains information regarding data collection, consent and statistics, while the following section builds up on the facts and examples presented thus far, and describes a project called Hero which is currently used actively to collect and process data from various systems and websites. In Section 6 we summarize our findings.

2. Education and Contests

Numerous countries and international organizations recognize education as a basic human right. Although teaching may occur in both formal and informal settings, primary and (sometimes) secondary education is compulsory in most countries. Additionally, online and electronic platforms enable students and professionals to learn at their own pace. Computer science, programming and software development are some of the most popular fields to study online, and newer video courses and tutorials usually encompass useful exercises that students can practice on.

Competitions are another major factor in education (Verhoeff, 1997). They can help to encourage students to perform better in school (for example, to earn scholarships), to form teams and study groups with individuals that have similar interests, or to win various awards and certificates. Some companies use competitions to hire the best professionals or university students, while schools or accredited organizations use competitions to increase interest in a certain field or subject. National competitions are used to select students that will later represent the country at global competitions (such as, for example, the International Olympiad in Informatics).

In the list below, we present some of the most popular competitions in mathematics and informatics. In the next section we will present several websites and grading systems which are used to award points to solutions of algorithmic problems. Some of these websites, like Codeforces, have blogs or pages that contain tutorials that students can use to learn programming languages, algorithms or data structures.

- **The International Olympiad in Informatics** (IOI, 1989) is a popular programming competition that takes place each year. Countries can send teams of 4 contestants, which compete by solving various algorithmic tasks. The IOI uses an automated grading system, and most of the tasks require students to write programs in one of the allowed programming languages.

- **Bebbras** (Bebras, 2004) is a competition that is organized in several countries, with the main goal of promoting informatics, computer science and computational
thinking among pupils. Most of the tasks are multiple-choice, fill in the blank, or interactive problems. Compared to the IOI, which has a very low number of female participants, Bebras is actually popular among both male and female students. The number of participating students increases every year (see Fig. 1). The competition is typically organized by school teachers, during regular school hours. Countries use different (custom) grading systems to organize the event, usually developed by the organizer or a partner.

- **The International Math Olympiad** (IMO, 1959) is an international competition in mathematics that takes place in a different country every year. The first event took place far back in 1959 in Romania. Countries send teams of up to six students to represent them at the prestigious IMO event. Google sponsored $1 million to the organization behind the International Math Olympiad in 2011.

In Macedonia, regional and national competitions are organized in several STEM-related subjects. For example, competitions in informatics are held in Macedonia since 1990, and they involve solving algorithmic tasks by coding solutions that are later automatically graded (Kostadinov et al., 2015). After several stages of competitions, the best primary and secondary school students represent the country and themselves at the Balkan Olympiad in Informatics (BOI), the International Olympiad in Informatics (IOI), and the Junior BOI (for younger students).

Starting recently, there is also a separate regional and national competition for primary school students in computational thinking, using Bebras-like tasks. This has led to an increase in the participation at the national competitions in informatics, and an improvement in the gender balance (due to more female participants). Awards are given to the best students from each grade (age group). Students that win a medal at an international competition, as well as their teachers, are awarded a monetary prize by the Ministry of Education of Macedonia.

![Fig. 1. Number of participants at Bebras.](image-url)
3. Grading and E-learning Systems

E-learning has a lot of advantages compared to traditional learning. In most cases, when schools or teachers discuss distance or e-learning, they talk about blended learning – or a combination of the traditional classroom with electronic technology. However, with the recent rise of popularity of Massive Open Online Courses (MOOCs), students and professionals are more and more likely to take courses completely through the Internet. A number of online platforms offer both free and paid courses in various areas: examples include Coursera (Coursera, 2012) and Udacity (Udacity, 2011). Videos, animations, documents, forums and quizzes are just some of the features that are offered by online courses.

E-learning doesn’t have to involve a traditional teacher with a degree in education. Some popular websites also offer courses created by professionals in different fields, so e-learning is used by students of various ages – from primary school students (where the school or teacher uses electronic technology in addition to traditional learning), to professionals (who are using online courses to acquire new skills). It’s believed that the rise of e-learning, the Internet and technology will help expand access to education to both the general public and businesses. Another big advantage compared to traditional learning is that online courses allow people to study at their own pace.

Software development is one of the fields where e-learning can help the most. Various platforms exist where people can learn by reading tutorials, watching videos, executing code and instantly viewing results, solving smaller tasks or running database queries with automated feedback and more. Competitions in informatics also use grading systems to automate the process of scoring solutions – much quicker and with less chances of error (a mistake can still be made if the grading system experiences issues, or if the tasks and their scoring criteria aren’t defined properly).

Table 1 shows examples of a few popular websites where software engineers go to learn, practice or compete. In this paper, we’re interested in all these types of systems, as we want to track how students learn or practice STEM.

Since the main topic of this paper is the collection of data and statistics, it’s important to understand what students and teachers do on these systems. Without the necessary

<table>
<thead>
<tr>
<th>Product</th>
<th>Type</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Udemy</td>
<td>Online learning platform</td>
<td>Video tutorials, discussions, quizzes, file sharing</td>
</tr>
<tr>
<td>Stack Overflow</td>
<td>QA site</td>
<td>Questions/answers, learning by viewing best answers on common questions</td>
</tr>
<tr>
<td>CMS</td>
<td>System for running programming contests</td>
<td>Grading solutions, answering questions, administering contests (self-hosted)</td>
</tr>
<tr>
<td>Codeforces</td>
<td>Site with contests &amp; tutorials</td>
<td>Contests, practicing in virtual competitions, learning from tutorials and posts</td>
</tr>
<tr>
<td>MENDO</td>
<td>Site with contests &amp; courses</td>
<td>Site with forum, wiki, national contests and courses for C++ and algorithms</td>
</tr>
</tbody>
</table>
information, it would be impossible to decide what data to gather, process or store. Additionally, even if we had the option of downloading everything, it would still be important to know how to merge that data with the other information from a user profile.

**Codeforces**, for example, allows users to participate in competitions organized on the platform. However, it also enables them to read tutorials on various algorithms and data structures, and practice by solving tasks given at previous competitions. When a user solves a task on his computer, he can send it to Codeforces, and the system will automatically grade it and show the verdict.

The platform contains hundreds of tasks (available in what’s called the “Problemset”), and all of them are marked with the appropriate tag referencing an algorithm, technique or a data structure (e.g., “binary search”). Students may practice by solving tasks at random, based on their difficulty, or based on a tag.

**Topcoder**, a similar site to Codeforces, also organizes competitions and provides a section where users can practice. Additionally, it has data science tutorials on various algorithms and data structures. Tasks are similarly marked with their type and difficulty. By looking at a user profile, it’s possible to see his rating and the latest competitions he participated in (i.e. his activity). It’s important to note that Topcoder hosts other types of competitions as well – design, development, bug races, and more.

On the other hand, the grading system used at the International Olympiad in Informatics, referred to as Contest Management System (CMS), is mostly used as a pure contest management system. Organizers will upload tasks to the system, and students will later participate in a contest by solving those tasks. Looking at a scoreboard, it is possible to see the rankings and the results of each contestant for each task.

The accredited organization in Macedonia for conducting competitions in informatics for primary and secondary school students on national level is the Computer Society of Macedonia. A competition management system called **MENDO** is used to teach students about programming and algorithms, and to prepare students for competitions in informatics (Kostadinov et al., 2010). In the last couple of years, MENDO has been used to organize almost all national competitions in informatics, as well as several Balkan and Junior Balkan Olympiads in Informatics. In total, more than 300 events have been organized, and more than 10,000 users have submitted around 350,000 solutions.

A forum, wiki and a commenting system allow users to ask questions, share solutions or discuss programming-related topics. A number of university courses have been organized on the platform by utilizing the various features and grading options.

Tasks with automated grading are a very powerful way to test the user’s knowledge and to provide an opportunity for practice. However, they sometimes influence students in a negative way – for example, a user being unable to send a solution because he is not following the system instructions (printing data in a wrong format, etc.). MENDO is an interactive system with various feedback features, one of which is the ability to discover issues like the one mentioned previously – i.e. it runs a solution-specific analysis and matches issues with a preprogrammed set of mistakes. Other options like the ability to download test cases, to view solutions, and to read open tutorials (with animations and interactivity), are also very useful and allow administrators to attract new participants without much involvement from their schools or teachers.
On MENDO, students also participate in competitions and other events (e.g., university courses). During a competition, participants can (optionally) receive feedback – depending on how an administrator wants to configure each task. After a competition or lab exercise has completed, students (and other visitors) can view the results of the competition. This means that, if we want to gather this data for a system that tracks or combines user activity, we can do that easily since the information is readily available.

It’s worth pointing out that there are many other systems and websites that software developers use. Some of them are used to organize events like Beaver – where students solve various interactive tasks. These are mostly open websites which allow students to view their results online. The Macedonian system (talent.mk) allows teachers and school principals to view all results, while students can only view their personal score (and download their certificate).

Some systems are developed and owned by companies, and are used to organize competitions such as Google Code Jam or Facebook Hacker Cup. Students and professionals participate in those competitions by registering on the event’s website. For Code Jam, for example, results are then published on the official website, but there are other platforms which track that data and publish it in a more readable format (for example, filtered by country, language or round). It is possible to view the relative rank of contestants versus others from the same country, view statistics about the programming languages that contestants use to solve tasks, and more.

Other systems have been developed by universities to organize their internal courses, quizzes or admission tests. These are mostly available via the Internet (in order to allow students to submit homework, and for teachers to view results or add tasks), but sometimes they might be behind a university firewall.
4. Data and Statistics

Data science and analytics are used by various products and websites to gain valuable knowledge about users and their behaviour. With the power of computer science, statistics and artificial intelligence, software solutions can describe models and predict risk or performance. Analytics are used in all fields of study – social sciences, legal, business, medicine and more.

For web applications, it’s common to use a premade solution like Google Analytics to both collect data, visualize it, and provide various other insights. Developers insert a small piece of code on a website, and the data instantly becomes available and visible on a dashboard. User and session counters, bounce rates, real-time reports, traffic sources, referrals, demographics and locations are just some of the information that can be easily tracked and analysed. Usually, other sources of data include server logs, the application’s database (relational or non-relational), hosting providers’ dashboards and APIs and more. Various free and commercial software solutions exist to parse this information, compare it side-by-side, and then visualize and group critical data.

It’s important to realize that users need to be informed on how their behaviour and data is tracked and stored. Privacy laws in countries and regions must be respected, and most websites already have a Privacy policy that describes what data is stored.

But, users are spending time on other websites which might be out of our control. Some of these websites have privacy policies which allow the sharing of data, and have APIs that other software systems can use to look and download information they might be interested in. It might be beneficial to track, store, aggregate and present this data for users or organizers, and thus help them make better decisions. A good example of this is Codeforces, which has an API that can be used to monitor what tasks users are solving, contest standings, problem statistics, user’s activity on contests and more. Each registered user on the platform has an API key which he can use to query information about him/her, but also info about other users and contest participants. Problem statistics can be used to discover easier and harder tasks, submission verdicts and resource consumption (like time and memory used).

The system used in Macedonia, MENDO, also has an API which is used to query and parse information. For each task, MENDO stores details about the number of users that attempted to solve the task, how many succeeded in doing so, details about the failed submissions (whether they were too slow or outputted wrong answers), and more. Additionally, for each user, the system knows which tasks were successfully solved, on which tasks the user failed, which competitions the user participated in, the lectures and tutorials that were completed and more. This data is mostly used by the students themselves when they connect to the website directly (outside of the API), as the system can tag (with a different color) each solved or failed task.

Most students who use MENDO like to see more information attached to each task or contest, as this helps them to make better decisions on how to spend their time on the system. Administrators constantly get requests for more such features. As an example, Fig. 3 presents some of the statistics and charts that are available after each
contest. Others not visible on the screenshot include the points distribution, number of students that solved each task correctly or partially, the speed of the top performers and more.

Of course, this data is also used by contest organizers to see how well their tasks were designed, and whether the point distribution is too low or too high.

Before proceeding to the next section and presenting our software, we will outline the two most common ways of retrieving data from other websites, and the issues that need to be considered.

The first option is to use the official API of the website. There are a lot of benefits to this approach: structured data, documentation, adhering to the site policy on what data can be retrieved, and more. Most websites allow the data to be consumed in either JSON or XML format, and most programming languages have libraries already available. Since Application Programming Interfaces are meant to be consumed and parsed by machines, the system will already have rate limits in place, and will expect those endpoints to be accessed by bots.

The second option is to use web scraping – i.e. to extract data from the website by parsing the HTML code. Since websites are designed for humans, this data might not be structured and thus may be difficult to parse. Also, some websites employ methods to prevent web scraping. The biggest advantage of this technique is the ability to access everything that a regular user has access to.
With both techniques, it’s very important to adhere to the site’s policy on which data can be retrieved and stored locally. If no such information is available, it’s common to ask the webmaster for permission and instructions. Although some webmasters might place restrictions on what load can be put on their server (so other users are not affected), most will allow scraping to occur if the platform is not a competitor product and you agree to link back to the original site.

After the data is transferred to a server, different database management systems can be used to store and analyse it. Depending on the size of the data, a relational or non-relational database can be used, or data can be stored with a cloud provider that offers storage services (and optionally, analytics).

Finally, it’s worth pointing out the importance of respecting user’s privacy – even when we store data which has been retrieved from other locations. Privacy is a major issue, and laws seem to be getting stricter by the year. If a person wants their data to be removed or hidden from a platform, that request should be fulfilled. With regards to programming and storing data about contest results and training activity, most software systems are open to users and they can then see their profile and use it to help improve (i.e. plan what to practice on next). Organizers can also use the data to track what students need to work on, and whether or not they are actively practicing and participating in competitions.

5. Software

In the previous sections, we outlined several systems that are used by contestants in algorithmic programming. Additionally, we presented a couple of ways in which the activity on those systems can be collected. The Hero app is a software application that tracks, stores and visualizes that data, and which is currently used by several organizers and contestants in Macedonia.

The initial idea for a software solution like this came 1) from the work that organizers needed to do in order to track contestants after the national cycle (in order to help them perform better at international competitions), 2) to make sure a student that did a bad result at one competition isn’t eliminated if his results at many other online systems are good, and 3) from various requests from contestants to learn more about the practice plans of students that performed better than them.

In the Hero app, contestants can be registered by configuring the usernames they use on various systems, or by searching for them by location (for example, contestant names on Codeforces can be listed by country). Additionally, it’s possible to view rankings by country or group. One example of grouping contestants is listing the activity of everyone who participated at the Macedonian Olympiad in Informatics (around 20 students). This in turn enables users of the application to quickly determine what contestants are doing.

Currently, the application tracks activity on the following systems:

- mendo.mk
- topcoder.com
- codeforces.com
As an example, when looking at a user named “John Doe”, it’s possible to see how he performed at every competition from the national cycle (since all of them are organized on mendo.mk), the tasks he solved by practicing on mendo.mk, TopCoder or Codeforces, or his rank/results during TopCoder or Codeforces rounds.

Of course, if one user has problems with a certain task, he can view which other friends from his group solved the same one, and ask for help to understand the solution overview or source code.

It’s important to realize that several other websites offer the option to see the activity of other users, and to organize or track a personal practice plan. However, Hero’s performance, grouping and visualization abilities separate it from other similar solutions. On Fig. 4 you can see a screenshot of the Hero app, showing student’s activity on the MENDO contest and training management system.

A rating value is calculated for each contestant, depending on his recent competition results. Exact numbers like the count of solved tasks are also readily available. The application has several additional views showing a calendar of future events, tutorials, compact views of profile data and more. These can be used by both organizers, teachers and contestants to plan practice rounds and to stay informed of various online events. New tutorials can be added by organizers and are grouped and tagged by their complexity. A calendar lists various online events on popular websites. This section of the

Fig. 4. Screenshot of the Hero app.
Platform for Analysing and Encouraging Student Activity on Contest and ... 95

platform is populated automatically by downloading event calendars available online. The main idea is to enable students and teachers to subscribe to these events, and get notifications on the beginning of each week, so they don’t miss an event they would otherwise be interested in participating in.

The implementation of the application uses both web scraping and APIs to get the necessary information. The fetching of data occurs on a schedule (every N minutes), in order to make sure the platform doesn’t put unnecessary load on the systems it connects to. Additionally, there is a special database table that stores the time when every call occurred, to guarantee that a programming mistake won’t lead to an unwanted denial of service attack.

From an implementation perspective, the Hero app runs on Node.js, a JavaScript runtime with a non-blocking I/O model that executes code server-side. Express.js is used as a web application framework, as it is fast and lightweight. This enables us to run the application on inexpensive virtual machines (as it’s currently the case) or dedicated servers, with minimal resource consumption.

Hero stores all data in a PostgreSQL database, and Docker is used to run both the backend application and the database. PostgreSQL is an object-relational database management system which can run both on Windows and Linux. This means that all technologies that are used by Hero are open-source and free. One of the best features of PostgreSQL is the JSONB column type, which ensures that all stored data in the column is valid according to JSON rules, but enables applications to store data that might have minor differences (for example, there is some data available for Codeforces contests that isn’t available for contests from MENDO). Hero attempts to store as much data as sent by outside APIs, even though some of that data might not be shown in the current version of the application.

Automatic backups to another server are scheduled every day, to guarantee that administrators won’t have to execute scripts to poll the same data that was already downloaded. Most database management systems have the ability to create backups without locking or disabling the database.

The technologies used on the frontend include React and Recharts. React is an open-source library for building user interfaces, maintained by Facebook and Instagram. Large web applications use it to show data and forms without reloading pages. The main purpose of Recharts, on the other hand, is to help developers to easily create charts on React. Charts are the perfect way to present data that can be quantified, and Recharts enables the creation of line charts, bar charts, pies, radar charts, tree maps, scatter plots and more.

Hero can only be accessed by authorized users (students, teachers or organizers). New users can be added by administrators through the application. Because authorization is needed, all profiles and combined data is private and only visible from inside the application.

Finally, it’s worth pointing out that every retrieval call (either via an API or scraping) is defined in a separate file. By creating an application using this method, it’s easy to add additional functionality in the future, and download data from more systems or websites. Every call is tagged with a system name or website url, and internal checks
common to all retrieval calls guarantee that no outside system will ever experience an undesirably high number of requests – even if a (new) developer makes a huge mistake when programming additional calls. Various middleware checks and database constraints are used by the application to restrict the storage of flawed data in the database, or the addition of duplicate data. For example, even if a user changes his username on an outside system, the application will try to match him or her by their e-mail address, full name, and past practice and contest experience.

6. Conclusion

Countries and various organizations are investing a lot of resources into STEM education, since those disciplines are needed to improve competitiveness and speed up development. Similarly, pupils and professionals interested in computer science are using various e-learning systems and websites to stay informed of new technologies, practice and improve their skills and get help when the need arises.

The rise of informatics has lead to the formation of several national and international competitions, like the International Olympiad in Informatics or ACM-ICPC. Students use various grading systems and websites to solve algorithmic tasks and prepare for those competitions. Most of these systems offer tutorials, videos, or automated grading, and are filled with useful data on how some students practice or what knowledge is needed to solve each task. This data can be collected, processed and analysed to help provide recommendations to specific students on how they can use their time more effectively when learning or practicing, and to inform teachers and organizers of any lack of activity. Privacy must be considered, as these systems help create user profiles and can easily visualize user activity.

Data collection, analytics and machine learning can be used in all fields – including medicine, social sciences, education and commerce. Multiple companies are investing in research associated with this area, and several open and commercial software solutions exist to facilitate the analysis of data and to provide recommendations.

Acknowledgement

The research presented in this paper is partly supported by the Faculty of Computer Science and Engineering, at Ss. Cyril and Methodius University in Skopje.
References


http://bebras.org/


TopCoder. https://topcoder.com


http://www.ioinformatics.org/locations/ioi97/ffutwrld/competit.pdf

B. Kostadinov is the founder of Cloud Solutions, an author, and a former competitive programmer. In 2014, he defended his MSc thesis in Intelligent information systems at the Faculty of Computer Science and Engineering, University “Ss. Cyril and Methodius”, in Skopje. He is one of the organizers of the national competitions in informatics in Macedonia, and the Beaver event.

M. Jovanov is an assistant professor at the Faculty of Computer Science and Engineering, University “Ss. Cyril and Methodius”, in Skopje. As the President of the Computer Society of Macedonia, he has actively participated in the organization and realization of the Macedonian national competitions and Olympiads in informatics since 2001. He has been a team leader for the Macedonian team at International Olympiads in Informatics since 2006, and member of the IOI International Committee since 2015. His research interests include algorithms, future web, and e-education.

E. Stankov is a teaching and research assistant at the Faculty of Computer Science and Engineering, University “Ss. Cyril and Methodius”, in Skopje. He is a member of the Executive Board of the Computer Society of Macedonia, and has actively participated in the organization and realization of the Macedonian national competitions and Olympiads in informatics since 2009. Currently he is a Ph.D. student at the Faculty of Computer Science and Engineering. His research includes analysis of program code correctness using different techniques, and its application to e-learning.