8th International Conference on Higher Education Advances (HEAd'22)
Universitat Politècnica de València, València, 2022
DOI: http://dx.doi.org/10.4995/HEAd22.2022.14195

Using data analysis to predict the students' trend of choosing preferred data storage

Georgi Cholakov, Asya Stoyanova-Doycheva
Department of Computer Systems, University of Plovdiv "Paisii Hilendarski", Bulgaria.

Abstract

Predicting trends is crucial for any business. No exception for education as well. Usually, this is a complex task that needs good planning and hard working to get to results. But sometimes luckily, a result from a study could be recognized as something that could reveal a potential trend, though it was not its primary goal, but spending some time digging deeper into data would pay off.

This paper presents results from a study of the students' trend for choosing favorite database type to learn and use, which was found during analyzing data from the software agents that work for our e-learning portal DeLC, serving as helpers for students and lecturers. These agents are there for very different purpose, but from the data they collect many interesting facts and behavioral patterns of our students could be revealed.

Keywords: *e-learning; software agents; data analysis; databases.*

1. Introduction

Our e-learning portal is called DeLC (Distributed e-Learning Center) and it's a project developed in Department of Computer Systems, University of Plovdiv "Paisii Hilendarski", Bulgaria, to serve our needs regarding distance learning, exams and other educational and organizational activities, subject of many studies as Stoyanov, S., Doychev, E., et al., (2012); Stoyanov, S., Zedan, et al., (2012). Among the most important advantages of this project is that its code base is available for researchers from that university, and thus possible are developing, reengineering, and improving most of the features it provides, allowing "insider" look of what's happening during this system's work, what data is stored, and the researchers are able to run many analytical processes trying to extract useful information and knowledge. Over the years, users from other universities started using it and thus it became a huge system, that combines functionalities and data from several satellite systems, which extend its capabilities. Among these extensions is IntelliDeLC, described in Cholakov (2013) – to provide a personalized e-learning environment with reactive and proactive behavior – in its nature this is agent-oriented extension, which provides environment that contains software agents. These agents are being developed and improved constantly and their functionalities, behavior, and latest results are discussed in many articles, among which: Cholakov (2020), Cholakov (2021), Cholakov & Stoyanova-Doycheva (2021).

The results of agents' work are subject to various analyzes, and they often reveal information, that is not visible at first sight. Such case was described in Cholakov (2021), when analysis was done on the results from automated tests assessments so far, made by the software agent that is dedicated to this task (EvaluatorAgent), for a particular course in our department – "Database management systems" (DBMS). The success/failure of the students was summarized by the points, earned during exams, and discovered which topics are difficult for

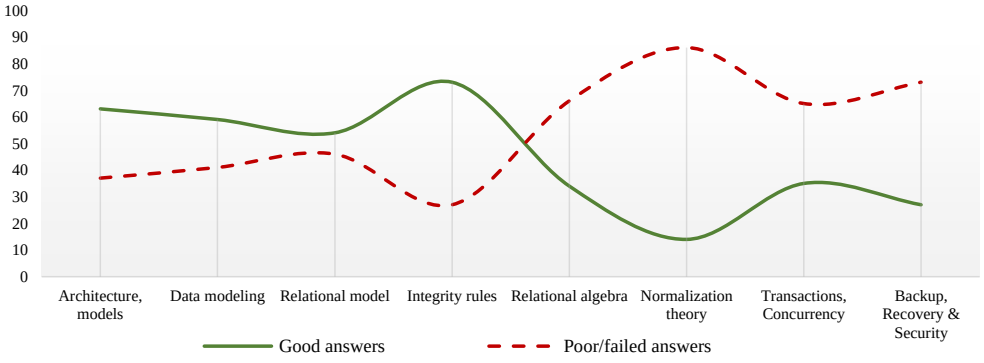

Figure 1. Percentage of the good/poor answers, grouped by topic for the DBMS course.

students. This is summarized and clearly visible on Figure 1 – when the materials are getting more theoretical and require more attention to be paid, the results are dropping significantly, and on the other hand, with more practical topics the results are higher.

It formed an interesting, but unwanted trend – students tend to rely on their practical knowledge to pass the exam, trying to avoid usually more complex theoretical topics (the assessment consists of two parts – practical and theoretical, which form the final grade, but the details will be skipped for brevity as they are not a subject of this study). Unwanted because it may lead to lower quality of knowledge and lack of analytical thinking. This information came as a side effect of our analyzes and led to course materials updates and other measures in order to keep students' attention, discussed in Cholakov (2021).

Now, could even more information be extracted from the data collected (Fig. 1) about students' behavior and preferences, and get to new conclusions?

2. The problem and results from further analysis

What happens to those students that ignore DBMS course and at a later stage they meet the requirement to use data storage of any kind for their projects in other courses or diploma thesis? It was worth to check out. A survey was conducted among the students and graduates, questioning what the preferred data storage type is – relational or NoSQL, used for their course projects and diploma theses, and then another trend was discovered. The survey was conducted online through Google Survey (Google Inc., 2022) and among the participants were 146 graduates and 274 undergraduates – all went through DBMS course. Here is the moment to mention, that our DBMS course educates primarily in relational databases. Many cases were observed when students use NoSQL databases just because:

- They have missed the relational databases course, don't know relational model at all or just don't feel comfortable with it.
- They find non-relational databases (particularly key-value and document storages) easier to learn and use, avoiding in-depth thoughts regarding relational theory.
- They want to keep up with the newest cutting-edge technologies – this is the smallest number of students, but they deserve special attention as they clearly see their future growth path and can distinguish between current market circumstances and what the demands would be in the nearest future.

And this was thought provoking – which are the main drivers for choosing the right database? Are our students lazy? Do they hate theory? Or are they bringing the future with all emerging technologies? Are NoSQL databases the choice of less educated? Or probably teaching DBMS should be revised with a different look, because the reality has changed. No simple or definite answer.

As Figure 2 depicts, among the students that achieved high results after the DBMS exam about 74% prefer using relational databases, the rest prefer other types; among the students with low results in DBMS exam the preference is split between relational and NoSQL databases; globally, the situation doesn't differ too much from the first part of the students, according to the most popular sources (DB-Engines Ranking (2022), Developer survey (2022), Top 10 Databases to Use in 2021 (2022)) – nearly 70% usage of relational databases.

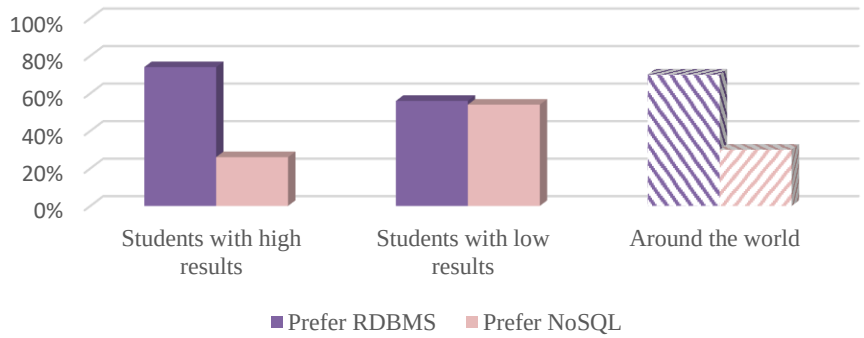

Figure 2. Percentage of students' preferable database type, and globally.

So, what those graphics above reveal – are the students lazy enough to push the progress further? Or probably relational model is not perspective to meet the modern world's needs? Both yes and no probably – it's a simple truth that:

- Practice makes perfect – learning new databases alone on demand trains problem solving skills, which is among most wanted qualities in IT branch these days.
- All things that one can do with NoSQL databases are possible with relational ones as well – so those who learned relational model didn't waste their time, as the statistics for global situation from Figure 2 clearly states (third data series).

The summary above should be considered as a red flag – the teachers need to review their understandings about databases, and this should involve updating courses, materials, and minds too – to reflect the reality properly. Some of these activities have already started – there is undergoing process of updating lectures and labs guides to include materials for modern databases, but it needs careful planning as the course still needs to fit in the same number of hours. To avoid this limitation, we already have an elective course for MongoDB (MongoDB (2022), the most popular among document-oriented ones, according to the statistics in the sites cited above) that students could enroll if they are interested in NoSQL databases, as it is a good starting point.

3. Conclusion

Using data analysis and data mining most of the times helps discover things that aren't obvious, and we witness their appliance in large range of domains – finance, education, medicine, social science, automotive industry, intelligent agriculture, to name a few. In our case revealed the trend of our next generation – what kind of database our students tend to use and what could be their career's choice. Discovering students' preferences helps us choosing the right tooling for education and gives direction for future updates of course materials.

Whilst investigating the most difficult parts in the teaching materials, we figured out that there are many students that don't try hard to learn the theory of relational model. Later, they tend to use non-relational databases as they find them easier to start with – and they do it quite well.

The results from this study could serve as direction what course materials must include to meet the modern requirements.

Acknowledgements

The research is supported by the project КП-06Н36/2 "Competition for financial support of basic research projects – 2019" of the National Scientific Fund of the Ministry of Education and Science in Bulgaria.

The statistics for global situation on database market are taken from the fellows from db-engines.com, towardsdatascience.com, and insights.stackoverflow.com, so special thanks to them for the interesting surveys they do each year.

References

Cholakov, G. (2013). Hybrid Architecture for Building Distributed Center for e-Learning. *PhD Thesis*. Plovdiv, Bulgaria.

Cholakov, G. (2020). Approbation of software agent Evaluator in a nonspecific environment for extension of its purpose. *2020 International Conference Automatics and Informatics (ICAI)*, (pp. pp. 1-5). Varna. doi:10.1109/ICAI50593.2020.9311346

Cholakov, G. (2021). Secondary effects converted to useful knowledge in e-Learning system. *International Conference Automatics and Informatics (ICAI)*, (pp. pp. 117-120). doi:10.1109/ICAI52893.2021.9639599

Cholakov, G., & Stoyanova-Doycheva, A. (2021). Model for Profiler Agent during unexpected educational circumstances. *7th International Conference on Higher Education Advances (HEAd'21).* Retrieved from http://ocs.editorial.upv.es/index.php/HEAD/HEAd21/paper/view/12841

DB-Engines Ranking. (2022, January 15). Retrieved from db-engines.com: https://db-engines.com/en/ranking

Developer survey. (2022, January 15). Retrieved from stackoverflow.com: https://insights.stackoverflow.com/survey/2021#overview

Google Inc. (2022, April 1). *Google Surveys*. Retrieved from https://surveys.google.com

MongoDB. (2022, January 15). Retrieved from www.mongodb.com

Stoyanov, S., Doychev, E., Valkanova, V., & Cholakov, G. (2012). Education Cluster for Intelligent Provision of eLearning Services. *DBKDA 2012: The Fourth International Conference on Advances in Databases, Knowledge, and Data Applications, Reunion, IARIA*, (pp. pp. 45-50).

Stoyanov, S., Zedan, H., Doychev, E., Valkanov, V., Popchev, I., Cholakov, G., & Sandalski, M. (2012). Intelligent Distributed eLearning Architecture. In *Intelligent Systems* (pp. pp. 185-218). InTech.

Top 10 Databases to Use in 2021. (2022, January 15). Retrieved from towards data science: https://towardsdatascience.com/top-10-databases-to-use-in-2021-d7e6a85402ba