Web application with Python and security of the information system

Petar Halachev
University of Chemical Technology and Metallurgy - Sofia, Bulgaria

Abstract: *The aim of the research is to develop a database management system for collecting, processing, storing and using information for the teaching of PhD students at a university using the high-level Python language.*
Studied and researched in the process of development are the main characteristics of the most widely used database management systems. The practical aspects of the design, creation and use of databases were analysed. Has been formulated the requirements to the functional capabilities of the developed database. For the development of the web-application was used Python programming language. The database model, the user interface and a set of reports were developed. A physical data model, oriented towards the design and the development of a database management system using the Python programming language was proposed. The main risks and threats to the security of information in the web-application are characterized. Guidelines for infrastructure protection are proposed.
Keywords*: WEB-APPLICATION, MODEL, DATABASES, PYTHON PROGRAMMING LANGUAGE, SECURITY*

1. Introduction

Information technology is one of the most important factors influencing the development of the society in the 21st century. The rapid development and penetration of information technology has led to the emergence of technical, social and legal phenomena. The most significant is the phenomenon of society's transition from traditional real-time interaction to the transmission of the information through electronic methods.

In the modern world, in various spheres of human life information technology is used. Each organization seeks to reduce the cost of time, material and labor resources in the process of its activities and to simplify the process of information processing. Through the usage of automated information processing systems and databases can be accomplished the solution of these tasks.

A number of researchers developed the conceptual foundations, the principles of the database design, the technology for their implementation and the systems for their management.

A. West and S. Prettyman [1] consider building interactive websites based on the MySql database. The emphasis is to install and launch a website with real applications in the shortest possible time.

K. Lang [2] designed the MySQL relational database management system and provided practical guidelines for installing and starting MySQL on a Linux based server. He shows how to use MySQL to create and manage databases in both command mode and batch mode, using SQL scripts. There are instructions for embedding MySQL to programs in different programming languages. The integration of MySQL with high-level programming languages for building dynamic web pages is considered.

J. Krogh [3] discusses the installation and setup of a MySQL connector for Python; connecting Python to MySQL; the configuration of the access to the database; the execution of SQL and NoSQL queries from a program written in Python; debugging and troubleshooting. He also considers the possibilities for storing data in different national languages using the support of the MySQL Unicode character set.

V. Siahaan and R. Sianipar [4] study Python-based software projects that use databases. The use of Python is suitable for designing and developing databases, as it contains libraries with rich functionality for opening, editing, adding new records and executing reports on various database management systems.

N. Chauhan, M. Singh, A. Verma, A. Parasher, G. Budhiraja [5] focus their research on the development of a database management system in colleges that other educational institutions could use. To develop this application was used Python. The information that is stored is accessible from anywhere in the educational institution. This system offers various features for students and staff members, which includes attendance and student ratings, available to both students and staff, but can only be updated by employees of a particular department. Students and staff have separate user profiles. The system includes a real-time library management subsystem, a college micro-transport management system and a dormitory accommodation management system.

Goals and objectives of the study

The goal set in the present study is the design, development and practical implementation of a database management system for collecting information about the training of doctoral students at a university with the application of a high-level Python language. In order to achieve the goal set in the project, it is necessary to solve the following tasks:

-To study and investigate the characteristics of different types of database management systems

-To analyze the practical aspects of the design, creation and use of databases;

-To study the state register of the documents for obtaining the scientific degree "Doctor" and the methodological recommendations for keeping and filling in the register;

-To conduct an analysis of the subject area and to formulate requirements for the functional capabilities of the developed database;

-To design a database of the trained and defended doctoral students at a specific university;

-To develop an application integrated with databases

-To collect and enter the necessary information;

-To implement the database using a high-level programming language Python - development of the database model, user interface, reference set, etc.

The approaches used are management of database systems and application of modern programming languages, in particular - a high-level programming language Python.

2. Research methodology

The report applies research methods related to solving specific tasks and achieving the goal:

-collection and statistical processing of information;

-study and analysis of normative documents and state standards regarding the information for trainees and defended doctoral students;

-conceptual modeling of databases and applications (MySQLServer) in order to build a system for information management in an educational organization;

-application of modern programming technologies (Python) in the design and creation of database management systems.

The Dutch programmer Guido van Rossum developed and created Python [6]. They call the Python programmers jokingly "Lifelong project dictators", which means that Guido monitors all language changes and makes the final decision to implement certain features when the situation requires it. V. Rosum also participated in the development of the educational programming language ABC. He then won the prestigious Free Software Award in 2001 while working for Google. Now the creator of Python works in Dropbox, which relates them to the cloud services.

According to K. Srinath [7], Python is a powerful, object-oriented programming language. He sets out the reasons for the accreditation of Python as the fastest growing programming language in recent times, backed by research on articles published in various magazines and popular websites. He presents the characteristics and the most important features of the Python

language, the types of database management systems supported by Python, its users and applications.

Python is object-oriented (works with fields and methods), has cross-platform compatibility, and can be programmed with the same set of features under Windows as well as MacOS, linux, * nix, and other popular operating systems. The program can analyze its structure and change it while the code is running. It is possible to execute instructions directly. It has functions for symbolic data processing. The language is aspect-oriented – divides the program into aspect-modules. This is a horizontal programming paradigm where can be added behavior (or function) to several classes that do not share the same vertically object-oriented inheritance.

Python has a minimalistic syntax, but it is no lower level and sometimes outperforms larger programmable environments. Minimalism allows you to increase the speed of writing programs, as well as increase the speed of reading code. The standard library of language modules includes an ever-growing set of different functions, and the user can easily add missing or new functions.

The reason for the popularity of the language is the clear and simply structured code.

Visual highlighting with spaces replaces the lack of bulky constructions denoting new classes, methods or threads. In such circumstances, it is much easier to track the progress of program development, it is easier to eliminate errors and add functionality. Structured Python code is much easier to understand, both for a beginner in programming and for a specialist who is in the early stages of using the programming language.

Python is multi-paradigmatic - programs are in one language, but in different styles. As with all programming languages, programs are using special rules - syntax.

Zen of Python includes a set of 19 rules for Python programming: simplicity is better than complex; complex is better than complicated; if the implementation is difficult to explain - the idea is bad; if the implementation is easy - maybe it's a good idea, etc. Programs written in Python should be simple to execute and easy to build, but if the situation requires it, the programmer is free to decide how to compile the code for it.

PEP8 is a set of general rules for writing Python code. It consists of code design recommendations, general tips and frequently asked questions with development examples. Although it can help solve the basic tasks of a novice programmer, in most cases teams of professional programmers complement PEP8 themselves, increasing the productivity of each team member as a whole. PEP8 defines mandatory rules when writing code:

-To use four spacing intervals, not to use tabs, not to mix intervals with tabs.

-The maximum line length is 79 characters. Must be used a slash ('\' character) and spacing of the new line to carry the line.

-Two blank lines must separate the first level functions and class definitions.

-Every import must be on a new line.

-Avoid additional spaces in parentheses, before and after commas, colon numbers.

-Frequently update the code comments.

-The comments should be in English.

-Avoid one-letter identifiers.

-Variables to have detailed and descriptive names.

-Do not compare Boolean variables with True or False, they should be evaluated directly.

Following the basic rules of writing Python allows you to create code that is equally easy to read and analyze. This writing approach is one of the main advantages of Python over other programming languages.

Based on the concept that a large amount of information must be organized in databases to show effectively the changes in the real world and to meet the information needs of the user, are based the main directions of the modern information technology. Databases are developed and operated under the control of special software systems called database management systems.

The different components, features and aspects of the database function relates to their variety.

These features include the nature of the information stored, the method of data storage, the structure and organization of the data, the method of accessing the data, the scope of the users.

Remotely performed is the management of the database - its creation, maintenance and the configuration of the user access by using special software tools - database management systems. Their range is very wide - when there is a need to store relevant and reliable information and quick access to it, including information systems, are used databases. Issues related to the maintenance of reliable and complete information and, in parallel, the timely registration of new documents are always relevant.

One of the key areas in the automation of the educational institution using information technology is the development of relational databases that can solve the problem of storing and systematizing information according to the specific requirements of the educational institution.

The problems of creating and designing databases and systems for their management are one of the important issues in the process of information technology development. Database management systems are evolving from single-user, which operate on a single personal computer, at a later stage as multi-user - based on the file server architecture, and then - on the basis of client-server architectures and management systems of distributed databases operating within global networks.

MySQL is widely used due to its advantages: it is easy to interact; with a wide range of programming languages; has high productivity; the software is open source; supports stored procedures and triggers. Along with these advantages, MySQL has some disadvantages: it is difficult to scale; MySQL does not work well with large amounts of data; does not fully comply with the existing SQL standard [8].

3. Database development

The developed database for doctoral students at a university targets to automate the processes of collecting, processing, using and storing the necessary information for the trainees and defending doctoral students, the issued documents. The design of the system allows increasing of the productivity of an educational institution by automating the processing and storage of information.

Before the stage of designing and developing the database for doctoral students at the university, the main theoretical aspects of the databases are studied, namely the characteristics, design, their classification and management systems.

Developed was a user interface of the application, and the data is then accessed with a convenient client application.

The developed database for doctoral students helps to solve problems related to increasing the productivity of an educational institution by minimizing the time for information processing. The information system serves to systematize the information about doctoral students at the university by automating this process. It also includes the work of extracting information on the status of the doctoral student, quickly filling in forms with the request of the employer for the authenticity of the diploma, automatic generation of reports for the doctoral students and elimination of errors when entering data.

A Python application links to a MySQL database by the MySQLdb library.

In the general case, you must additionally install it.

To execute an SQL query (SELECT, UPDATE, INSERT) for working with data, the following construction is used:

First, the connect function is called, and information about the MySQL data server is given as parameters - username, password, database name. Then you create a cursor to execute the SQL queries. After executing the queries, you process the result with an iterative construction for, which prints the first column of the result on the screen.

To close the connection we call the close function (Table 1):

Table 1. *Executing an arbitrary SQL query to the database*

```
#!/usr/bin/python
import MySQLdb

db = MySQLdb.connect(host = "localhost",
                     # usually localhost
                     user = "user",
                     # username
                     passwd = "*****",
                     # password
                     db = "PHD_STUDENTS")
                     # name of the database

# Creation of a Cursor Object
# Allows the execution of SQL quires
cur = db.cursor()

# Execution of a SQL query
cur.execute("SELECT * FROM DOCTORANTS")

# Printing the first column of the result
for row in cur.fetchall():
    print row[0]

# Closing the connection
db.close()
```

During the development of the database, was selected the Relational model. The model is on Fig. 1:

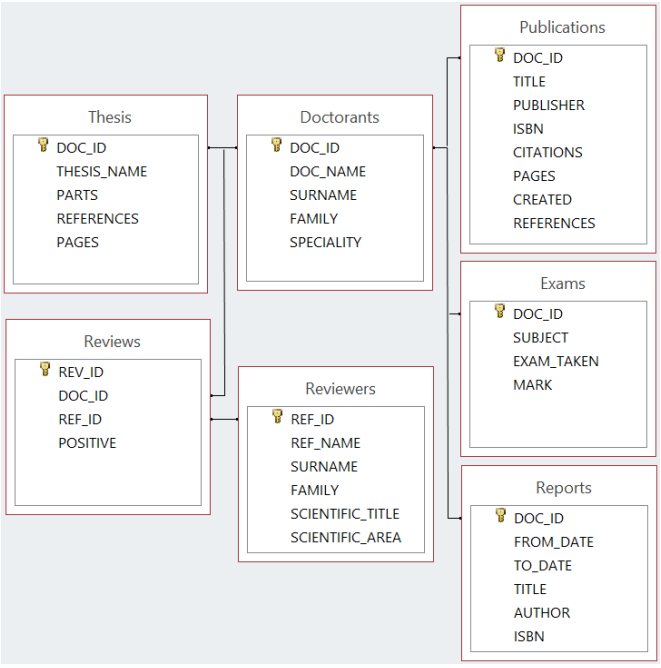

Fig. 1. *Relational database model*

4. Security of the application

The HTTPS protocol relies on TLS (transport layer security) to provide a secure channel for communication between two endpoints [9]. The means of symmetrical encryption using a temporary encryption key for the session protects the communication over HTTPS. To exchange the temporary key is used asymmetric encryption. To enable the usage of a SSL protected connection was installed a SSL certificate.

Some of the main threats to the information security of users in a web environment are user authorization, data encryption, physical security in the operation of equipment, harmful traffic:

-User authorization - When using a web services, the most common way to authorize users is the password protection.

The application of tools such as certificates (HTTPS) and electronic signatures provides greater reliability to users.

-Separation of user data - A problem with the web technologies is the separation of the data and applications from the users. The best option is when each user uses an individual virtual machine (VM) and a virtual network VPN (Virtual Private Network).

-Data Encryption - must be encrypted the connection channels and data stored on the web server. There is usually no problem when exchanging data over the network, as the secure HTTPS protocol provides the access to the data. However, when controlling data on the web, there are problems with the usage of encryption of one key for all accounts, and the hacker gaining access to the key could gain access to all the data.

-Physical security - Some measures that would prevent the possibility of leaking information about customers to foreign or bribed associates include: biometric identification, smart cards, video surveillance, regular checking of access logs and more.

-Harmful traffic - Deploying applications to the web server complicates the task of ensuring traffic safety between applications and the virtual machine. If hackers find a vulnerability in one virtual machine, they can easily access the others.

Carrying personal belongings (Bring Your Own Device) The trend of BYOD (Bring Your Own Device) [10] is increasing - carrying personal smartphones, tablets, laptops, etc. from employees to work, which leads to increased risks to security of the information of the organization.

Malicious participants often use the main security vulnerabilities of web applications:

-Cross-site scripting (XSS) protection - XSS attacks allow a user to inject client side scripts into the browsers of other users.

-Cross-site request forgery (CSRF) protection - CSRF attacks allow a malicious user to execute actions using the credentials of another user without that user's knowledge or consent

-SQL injection protection - SQL injection is a type of attack where a malicious user is able to execute arbitrary SQL code on a database. This can result in deleted records or data leakage.

-Clickjacking protection - Clickjacking is a type of attack where a malicious site wraps another site in a frame. This attack can result in tricking an unsuspecting user to perform unintended actions on the target site.

-SSL/HTTPS - it is always better for security to deploy a site behind HTTPS. Without this, it is possible for malicious network users to sniff authentication credentials or any other information transferred between client and server, and in some cases – active network attackers – to alter the data sent in either direction.

-Host header validation - must be used the host header provided by the client to construct the URLs in certain cases. The sanitization of these values prevents Cross Site Scripting attacks. For Cross-Site Request Forgery, cache poisoning attacks, and poisoning links in emails can be used a fake Host value.

-Referrer policy - Browsers use the Referer header as a way to send information to a site about how users got there.

-User-uploaded content - If your site accepts file uploads, it is strongly advised that you limit these uploads in your Web server configuration to a reasonable size in order to prevent denial of service (DOS) attacks.

-Additional security topics - While Python provides good security protection out of the box, it is still important to properly deploy your application and take advantage of the security protection of the Web server, operating system and other components. [11].

Conclusion

One of the effective ways to improve the quality of training is its automation with the help of modern computer technology, namely with the application of databases and software applications. In this way, it is possible to speed up the process of information processing significantly, to extract in a timely manner appropriate and reliable information about the PhD students and to compile reports according to set criteria.

The software application allows entering, editing, viewing, storing and deleting information about university doctoral students. The developed database facilitates the preparation of various types of reports and speeds up the process of obtaining the requested information and making management decisions. The implementation of the software application is aimed at improving the work of university staff, doctoral supervisors, optimizing the work with data and ensuring their reliable storage.

The practical orientation of the program application is you can apply it in other educational institutions with a similar profile.

5. References

1.Practical PHP 7, MySQL 8, and MariaDB Website Databases: A Simplified Approach to Developing Database-Driven Websites, September 2018, Apress901 Grayson Street Suite 204 Berkely, CA, Adrian W. West, Steve Prettyman

2. MySQL Database System, 28 August 2018, ISBN 978-3-319-92429-8, K. C. Lang

3. MySQL Connector/Python Revealed, ISBN 978-1-4842-3693-2, Jesper Wisborg Krogh

4. Learn SQLite with Python: Building Database-Driven Desktop Projects, Sparta Publishing, Sep 29, 2019, Vivian Siahaan, Rismon Hasiholan Sianipar

5. Implementation of database using python flask framework, 20 December 2019, Nidhi Chauhan, Mandeep Singh, Ayushi Verma, Aashwaath Parasher, Gaurav Budhiraja

6. https://www.python.org/ - Python

7. Python – The Fastest Growing Programming Language, International Research Journal of Engineering and Technology (IRJET), Dec-2017, K. R. Srinath

8. https://www.mysql.com/ - MySql

9. Rolf Oppliger, Security Technologies for the World Wide Web, 2003 Artech House Inc. ISBN 1-58053-348-5

10. http://newhorizons.bg/blog/2014/01/information-security-threats-2014

11. https://docs.djangoproject.com/en/3.1/topics/security/