An investigation into the design and development of a Python based data processing and graphical analysis platform

Pu Zhao

Department of Life Sciences, Imperial College London, Egham, SW7 2AZ, The United Kingdom.

Abstract: In the context of the wide application of big data, cloud computing, Internet of Things and other advanced technologies, the design and research of data processing performance and graphical analysis platform based on Python programming language has become a new direction and new goal for the innovative research and development of data analysis platform. In order to enrich the relevant research theories and put forward new ideas and paths for the design and development of graphical analysis platforms, this paper is based on the data processing of Python, and adopts the methods of reverse analysis and subjective assessment to explore the design and development of a new type of graphical analysis platforms with the core of the Python web crawler technology, with a view to providing a basis for the data processing and graphical analysis platforms to quickly analyze and process the relevant data and information and designing and developing a new graphical analysis platform with a visualization and analysis system. Provide the basis for the design and development of a new type of data processing and graphical analysis platform with visual display performance, intuitive and efficient presentation of the pattern of change and trend of data, to provide a reference basis for related research.

Keywords: Python; Data Processing; Graphical Analysis Platform; Design and Development

1. Introduction

In recent years, Python technology is widely used, this new intelligent data processing and analysis can be achieved through the visual display of data graphics, the warrior of massive data analysis results, graphical analysis in a scale and visualization of data management to deal with different types of data information. In this regard, the design and research of data processing and graphical analysis platform based on Python has become a popular topic focused on by experts and scholars at home and abroad. Compared with the traditional Excel data storage and processing methods, a single icon, pivot table and formula processing and analysis can only deal with the relationship between simple data information, it is difficult to do effective processing of complex data, there are certain limitations, is not applicable to a number of industries rapid processing of data information and graphical analysis of the needs. Therefore, the study uses Python, a lightweight programming language with a high degree of flexibility, to analyze the design and development of the data processing and graphical analysis platform based on Python, in order to build a new platform with integrated functions such as visual display, rapid analysis of data, import, storage, cleanup and management, to build a scientifically sound database, to set up the data input and output channels, to output the relevant data in a graphical analysis way, and to take the textual analysis of data. It is designed to construct a scientific and perfect database, set up data input and output channels, output relevant data by means of graphical analysis, and output problems by means of text, so as to provide a basis for the effective processing and utilization of data and information.

2. Overview of Python Web Crawling Technology

Python is the mainstream dynamic scripting language today, presented as a flexible programming language that can be used to automatically browse the Internet and automatically collect and acquire the required data programs from web pages. Web crawler selects the URL, initiates HTTP request, parses the server response content, collects useful data and new links, stores the collected data, and adds the newly acquired links to the queue and waits for further crawling, and stops data crawling after all the links are crawled or after a certain depth of crawling in a cyclic operation, and the flow of the Python technology is shown in Figure 1. The development of the crawler mainly relies on the simple syntax of the Python language, rich database and good adaptability to the flexible development of data analysis and processing. Currently, the commonly used Python tools and libraries are Requests library, which can simplify HTTP requests, while Beauti fulSoup can be used to parse XML and HTML; Scrapy framework is suitable for processing large-scale crawling projects; Selenium can accurately simu-

late the browser behavior [1]. In addition, the combination of CasperJS and PhantomJS with Python can improve the processing efficiency and quality level of JavaScript rendering web content.
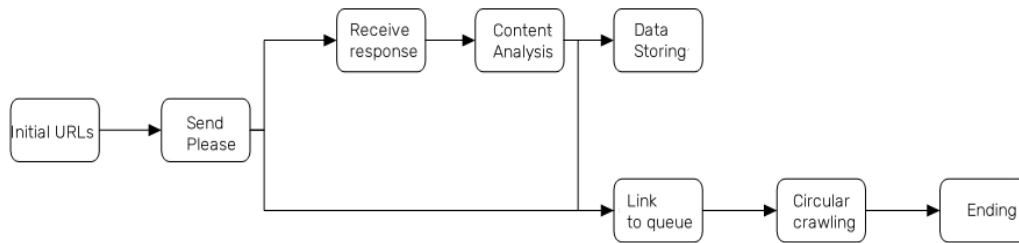

Fig. 1 Python web crawler workflow

3. Python-based graphical analysis platform design

3.1 Overall design ideas

The core of the design and development of the Python-based graphical analysis platform for data processing is to use reverse analysis technology, take the reverse analysis and competent evaluation of the design, and unify the file formats generated by different types of data parsing software. Provide convenience for users to use data processing and graphical analysis platform, and improve work efficiency under the condition of no file format switching. Figure 2 shows the design diagram of the technical equipment in the application of Python-based data processing and graphical analysis platform, including 9 functional modules respectively.
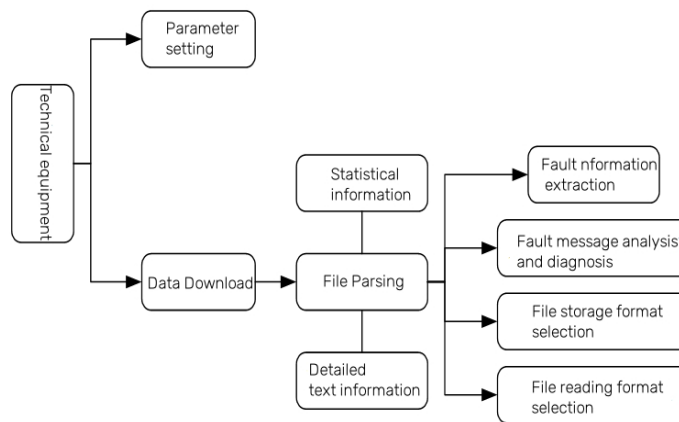

Fig. 2 Design of Python graphical analysis platform

3.2 Platform core technology

This study utilizes the open source Python2.7 language program to write visual data analysis software, while choosing os, sqlite3, json, codes, xlrd, wx and matplotlib modules, etc. Python2.7 can be downloaded through the Internet, and various types of functional modules are downloaded using pip. In order to realize the competent evaluation, to take the questionnaire survey, effective processing of questionnaire survey results can reflect the advantages of visual data analysis software application features, the results of the survey will be stored in the database, classification and integration of different data types, through the data targeted processing and analysis of the data results in the form of electronic documents, pictures and other forms of display to optimize the performance of the data processing and picture formal analysis platform.

3.3 Platform architecture and input

In the platform architecture and input design, the graphical presentation of data to achieve subjective assessment and visualization, but

also need to build a complete structure and scientific database. The structure of the database can be divided into two layers, the inner layer can be combined with basic information, such as personnel, questionnaires and questions and other information, the inner layer of information is formed by the initialization file, in which the personnel and questionnaire information is input by the questionnaire; the outer layer of information is the basic information of the device, for example, the car, including the model information. A certain number of testers are selected for subjective evaluation, evaluating and scoring the human-machine problems of a certain car model, setting the problems as whether the initial position of the brake pedal is appropriate or not and whether the steering wheel position is appropriate or not, etc.

In the evaluation of personnel information, it is mainly based on basic information such as name, gender, age, height, weight, driving age, etc., while the vehicle information is license plate number, car name and brand; the questionnaire information includes the content, type, and x and y coordinates of the question on the corresponding picture of the car; the questionnaire information includes personnel information, car information and question scoring. The input structure of the platform should be composed of tables of people, cars, questionnaires and questions.

3.4 Visualization

Running the visualization software will pop up an operable graphical user interface in order to take a graphical form of analysis and display relevant data content. The interface includes operations such as database, car models, questions, questionnaires and results, importing questionnaires and questions from different car models into the database, conducting data out and graphical analysis based on Python web crawler technology, and updating the network database in real time. Among them, the content of the questionnaire should be combined with the car model to develop appropriate entries, and the same questionnaire can also be used to analyze the problems of different car models, in which the ergonomics problems are the height of the door sill, head space, leg space and the height of the door armrests, respectively.

Vehicle occupants have some differences in the way of problem evaluation and analysis, based on different standards. Therefore, in order to obtain more comprehensive and detailed data, it is necessary to collect evaluation data from more occupants by means of subjective evaluation. At present, the models are constantly enriched and the ergonomics investigation questions are constantly changed, which makes the data processing and graphical analysis platform increase the amount of data to be evaluated in the data evaluation. Therefore, the database constructed based on Python is utilized to compile SQL statements in the database and create a one-time procedure to realize arbitrary calls to achieve accurate evaluation of large-scale data and improve the speed of database execution. Meanwhile, in initializing the database, relevant tables are formed, and the Cars.xlsx and Questions.xlsx files are utilized to add the entries related to the tables Cars and Questions. Utilize the print database to print all the entries of the tables.

According to the graphical display of the table entries derived from the comprehensive analysis of the following:

(1) editing, increase the model: the model information will be entered, according to the number of input automatically select the edit and increase the function. In order to facilitate the call and view the model data has been deposited into the database, but also on the basis of the original data and information, add new data and information;

(2) Delete the model: the model information to be deleted into the input, the accurate results of the options, if the results are not looking for a pop-up box will be prompted. Which models, questionnaires and questions have the function of deleting data information options to facilitate the deletion of obsolete model information. Ensure that effective information is stored in the database, and at the same time ensure the efficiency of database operation, enhance the operability of data storage, and provide convenience for the solution of problems in graphical analysis and processing;

(3) Increase and Edit Problems: The corresponding problem information is inputted, and the function of increasing and editing problem information can be selected automatically, all problems correspond to the corresponding numbers, and through the input of problem numbers or information, different Through the input of question number or information, it can accurately find out the evaluation of different human-computer problems by different groups of people and show them with visualized graphs;

(4) Delete questions: the question number or information to be deleted is inputted into the analyzing platform, and the query results are obtained. In the deletion of problem data, it is necessary to carefully check whether the query results are the same as the input data, and careful operation can prevent the deletion of usable data and avoid the phenomenon of operational errors [2].

4. Graphical Analysis Platform Features and Advantages

The design and development of the Python based data processing and graphical analysis platform uses mixed programming of C# and MATLAB in the graphical display. The use of MATLAB compiler to compile *.m file, transformed into a dynamic link library file *.dll for C# call, through the hybrid programming, so that the combination of C# and VB visualization operations, and based on the high efficiency characteristics of the C + + + system operation, to improve the operational capabilities of the platform analysis, to support the Python network crawler technology and .NET convenient component programming for data processing and graphical NET convenient component programming for data processing and graphical analysis platform design and development. At the same time, taking into account the lack of statistical analysis, the platform design and development process, but also based on a simple and practical vision, will be *.dll file saved to the appropriate directory, *.dll that is, the dynamic link libraries available for .NET, through the use of .NET call *.dll file, to achieve the graphical display of the data graphical visualization of the data, all kinds of data information in the form of graphical analysis of the form of presentation.

Based on the Python design and development of data processing and graphical analysis platform technology features are obvious, can be read and write data and increase or decrease the way, add new subjective evaluation data, stored in the database, to facilitate the subsequent data problem query, assessment and analysis; and will be useless data deletion, and timely update the database content [3]. At the same time, the corresponding data can be used, through data processing and analysis, to graphically display the specific location of the human-computer problems to achieve visualization, but also based on the text output of the human-computer assessment, analysis of the corresponding problems to produce accurate results, especially for the processing of complex data, due to the complexity of the data relationship with the large volume of factors such as traditional data processing and graphical analysis platforms can not be extracted from the valid data and display them in a visual graphical way, but the Python-based graphical analysis platform can intuitively and effectively operate, observe, and compare and analyze the relevant data through the use of visual output interface, graphically displaying the internal laws and characteristics of the data information [4].

5. Conclusion

In summary, the design and development of data processing and graphical analysis platform should introduce new technologies and fully utilize the power of big data, the Internet and other advanced technologies. This study analyzes the content of data processing and graphical analysis platform design and development based on Python web crawler technology, fully considering the difficulty of analyzing the large amount of data and complex data relationships. Therefore, the research adopts the reverse analysis technology and subjective assessment method to design the visual analysis software for subjective assessment of man-machine engineering, and at the same time, it uses Python programming language to carry out the calculation processing and analysis of data information, constructs an intuitive and accurate visualization interface, and optimizes the performance of the data processing and graphical analysis platform by means of visual operation, through the display of intuitive graphical data, to realize the man-machine engineering The rapid inspection and analysis of the problem is realized by displaying intuitive graphical data.

References

[1] Wang Shun, Li Wei, Zhao Shiju, et al. Design and teaching application of precipitation titration learning software based on Python [J]. Chemistry Education (Chinese and English), 2023, 44 (18): 113-119

[2] Li Yun, Yang Shaoqin, Fang Yanjie, et al. Development of a Batch Mapping Tool for Plantation Forest Design Based on Python - Taking Chun'an County, Zhejiang Province as an Example [J]. Zhejiang Forestry Science and Technology, 2023, 43 (05): 108-111

[3] Sha Yongbing, Chen Feiming, Cao Deqin, et al. Construction of a Time Series Matching and Disposal Model for Main Alarm Information Rules of a Hydroelectric Control Platform Based on Python Data Visualization [J]. Hydroelectric Energy Science, 2023, 41 (05): 182-186

[4] Du Anliang. Development of CalculiX and Sap2000 Preprocessing Interface Based on Python Language [J]. Architecture, 2022, 52 (S1): 446-449