

Construction and Research of College Student Certificate Information Service Platform

Wenli Ren^{1,a,*}, Jie Tang^{1,b}

¹School of Information, Shanxi University of Finance and Economics, Taiyuan, China

^a 643949600@qq.com, ^b 2449177808@qq.com

*Corresponding author

Keywords: Big data, professional qualification certificate, web crawler

Abstract: With the increasingly serious employment situation, it has become a major trend for college students to choose relevant certificates during their school years. Relevant research shows that as a mainstream group that uses technology to solve problems, it is often difficult to obtain the required information quickly and efficiently. At the same time, under the extensive certificate information and the exaggerated propaganda of the research institutions, it is more difficult for contemporary college students to define their own goals. The advent of the era of big data makes the massive acquisition of network information easier. The development of data analysis technology makes the analysis and processing of network information more convenient.

1. Research on thermal phenomena

According to relevant research, 34.34% of the students believe that the research can learn more knowledge and enhance their own value; 57.63% of the students believe that the research can help the employment. Faced with the increase of employment pressure, more and more college students choose to improve their competitiveness through research and verification. The phenomenon of "researching hot" has become the norm in the university campus. According to the survey, nearly 90% of students are preparing for research or have already taken the test. However, this is enough to illustrate the huge demand for certificates by contemporary college students.

At the national level, the "National Vocational Education Reform Implementation Plan" promulgated by the State Council clearly stipulates that the vocational education in the new era will be further improved, and the research talents will be vigorously supported, which will prompt the contemporary college students to set off a "research hot", but most students do not have the knowledge of the research information. Accurate, the search process is too inefficient, because the information on various types of certificates on the Internet is too complicated. Various forums and post bars have different opinions on the gold content and demand of a certain certificate, and the excessive publicity of training institutions makes college students Difficult to question whether to test or test the certificate.

2. The development of big data

Big data is not an exact concept. It is more of a way. Big data is something that people can do on the basis of large-scale data. These things cannot be done on the basis of small-scale data. Big data. It's a new type of ability to analyze data in ways that have never been seen before, and to get products and services of great value. The era of big data is the result of the combination of massive data and computer capabilities. To be exact, the data we leave on the network and the large amount of data generated by the Internet. Big data can collect, store, clean, and go through these advanced technologies. With heavy and reasonable calculations and analysis, the era of big data has opened up another era in which human society uses data values.

3. Platform concept

This platform takes the problem of solving the information search and course selection of contemporary college students as the starting point, and uses Python crawler and big data analysis technology to obtain relevant information of various certificates and evaluation information of training institutions or learning platforms, and cleans the data. De-weighting, standardization and targeted analysis, providing users with information such as industry recognition, market demand, difficulty in examination, and institutional evaluation, saving time and cost for users, improving information acquisition efficiency, and the platform will complete the verification calendar and certificate navigation. The establishment of functions such as interest selection, institutional scoring, and course optimization provides a “one-stop” service for user research.

4. Technical analysis

This platform is an integrated service platform. It is mainly composed of mobile client and desktop webpage. The development of mobile platform mainly adopts the current mainstream development framework. The webpage adopts responsive webpage design, which is convenient for mobile phone and desktop browsing to reach the mobile terminal and the desktop end in the website experience. The platform operation will adopt cloud computing solutions, cooperate with database storage services, and adopt high concurrency framework in the design to optimize for multi-user access. The operation content will cover the certificate information inquiry, the timetable of the research, the recommendation of relevant institutions or courses, and the exchange of netizens and forums.

Big data analysis is developed in Python and uses the pandas data processing framework. Pandas is a NumPy-based tool that can handle many different data types. When using pandas for processing, be careful to clean the data to avoid data interference. For example, the default `dropna()` method discards all rows with missing values. You can also specify the upper limit of the discarded data. Pandas will discard the rows with more missing values. The `map()` method can be used to generate new data according to the specified mapping relationship based on the existing data. The value is also replaced and the data is discretized during cleaning.

5. Access to and integration of certificate information

The arrival of the era of big data has provided strong support for the acquisition of information on various professional qualification certificates. Making full use of this emerging technology can solve a major problem for college students.

The crawling of the certificate information is developed in the python language, and the crawling operation program is written according to the content of the website. However, due to the variety of types of websites, different crawlers should be written according to the types of websites. For ordinary static web pages, use the URL link get function that comes with ordinary python to crawl. For dynamic websites, because the dynamic website loading often uses JavaScript, the ordinary post crawling using static web pages is invalid, and the phantomJS emulation browser should be used for data crawling. At the same time, many websites are designed with anti-crawling mechanisms to block the source IP of the crawl. IP proxy technology should be used to prevent IP from being banned. For the case where the proxy IP is often forbidden, the `scrapy-crawlera` package can be used for crawling.

In addition to the use of crawling information, there are also data obtained through cooperation with relevant institutions or platforms. For these large amounts of data, preliminary data analysis will be used for classification. The Kmeans unsupervised learning clustering algorithm can be used for automatic preliminary classification and the results can be processed in combination with data visualization. Visualization technologies include: tree-filled treemaps, topology layout techniques, parallel coordinate system visualization, and data overlay visualization.

6. Platform construction

The platform is divided into a desktop webpage and a mobile application, and the mobile application is further divided into an Android application and an Apple application. In the construction of the desktop website, the back-end server and database + front-end bootstrap responsive web design are used for construction. These technologies are mostly ready-made frameworks, which are convenient for rapid development and have high feasibility.

In the database aspect, SQL Server can be used to store metadata in the early stage. As the amount of video and the number of users increase, data can be divided into multiple partitions, the server can be expanded, the database can be divided into multiple shards, and different users can be assigned to different shards. Spread and read the database.

The front-end web page can use JavaScript and other technologies, and adopts a separate mode at the front and the back to facilitate separate maintenance and development of the front end.

The Android application can use the Application Framework development framework. Its system architecture consists of Linux Kernel, Android Runtime, Libraries, Application Framework, and Applications.

The Linux Kernel acts as an abstraction layer between hardware and software, hiding specific hardware details and providing a unified service to the upper layers.

Android Runtime can be understood as an instance in the Dalvik virtual machine. The Dalvik virtual machine is based on registers and provides basic functionality by the Linux kernel.

Libraries refer to a collection of C/C++ libraries included with Android for use by various components of the Android system. These features are exposed to developers through the Android application framework. Easy to use, development is extremely feasible

Application Framework

By providing an open development platform, Android enables developers to build extremely rich and innovative applications. Developers are free to take advantage of device hardware, access location information, run background services, set alarms, add notifications to the status bar, and more. Developers can fully use the framework APIs used by the core application.

Applications assemble a core application collection for Android to enable a variety of application development, and an open source-based development framework is also beneficial for project development.

Apple application development using Apple's xcode tool, you need to use Objective-C language to write, Apple-side application development is more convenient, the main use of the framework is Foundation and UIKit.

The Foundation framework provides basic system services for all applications, creating and managing collections, accessing images and other resources stored in applications, creating and managing strings, submitting and receiving notifications, and more.

The UIKit framework provides the ability to create touch-based user interface-based classes. All iOS apps are based on UIKit, and UIKit provides a mechanism for drawing on the screen. UIKit also builds and manages your user interface by capturing components displayed on the screen and captures touch and movement-based events, rendering text and web content, as well as creating custom user interface elements.

Core Data, Core Graphics, Core Animation, and OpenGL ES frameworks also need to be used in development.

The Core Data framework manages the application data model.

Core Data provides object management. With Core Data, you can create model objects and manage them. You manage the connections between these objects and modify the data. Core Data provides built-in SQLite technology to efficiently manage data.

The Core Graphics framework helps you create graphics.

Core Animation can create advanced animations and virtual effects, as well as create custom animations.

The above technologies are all off-the-shelf technologies, which are more convenient to develop and have higher feasibility.

7. Implementation feasibility

In the platform development, the server side can use the server computing platform provided by Alibaba Cloud or Tencent Cloud. The project can be developed online and then deployed to an online server for testing. The Android client and iOS client organize the development team to develop separately and test before going online. At the same time, for subsequent maintenance, software updates are performed in the form of iterative release.

8. Analysis of social benefits and economic benefits

8.1. Economic benefit

We need to consider the cash income and non-cash income of the development website. In terms of cash income, we must make reasonable estimates, try to increase cash income, and avoid the website development rate of return is too low; in terms of non-cash income, we need to comprehensively evaluate the website's functional requirements, quality requirements, development direction, etc. when predicting the development cost of the website. By evaluating the labor input, you can basically calculate the development cost of the website. As a traffic port, once the traffic is available, it will have certain benefits. After the number of users reaches a certain level, the platform will use advertising, user traffic and website transaction rebates as the foothold of profit, form a "three-in-one" profit model, and continue to explore profit points in order to maximize the benefits. Through cooperation with relevant training institutions, the company launches its audition course on the website platform, and charges a certain fee to it as part of the profit point, which not only satisfies the marketing purpose of the merchant, but also satisfies the learning needs of the user.

8.2. Social benefit

Colleges and universities are an important educational base for the formation of college students' reserve knowledge and ability. College students receive higher education, learn and master professional knowledge, establish a correct concept of entrepreneurship, and lay a certain foundation for the preparation of entrepreneurial technology. In the context of the rapid development of Internet technology, we should make good use of the professional knowledge we have learned to enhance our entrepreneurial awareness, and lay the foundation for us to use various advanced technologies and successfully enter the society in the future.

In recent years, the number of university graduates has increased sharply, the employment situation of college students has been severe, and the phenomenon of "researching hot" of college students has continued to heat up. The "judging families" of colleges and universities are everywhere. The appearance of the "research hot" phenomenon is not an accidental factor, but a result of the comprehensive effects of universities, college students, and society. Under the new situation of "Internet +", "Easy Testimony" website meets the needs of the market, provides professional information for all kinds of certificates and relevant guidance for research, and is a one-stop comprehensive service platform for college students to test and test, and realizes certificate information. Searching for a shift from a needle in a haystack to a precise search, helping the university to rationally select and improve its career planning and competitiveness, and become a highly talented person in the society.

References

- [1] Wang Jianglong, Wang Xiaohong. Implementation based on Python crawler technology [J]. Computer programming skills and maintenance, 2019 (09): 18-20+41.
- [2] Wu Yongcong. On the crawling and analysis of webpage data of Python crawler technology [J]. Computer Age, 2019 (08): 94-96.
- [3] Zhang Qiang, Qin Yadong, Zhao Baolin. Analysis of the phenomenon of "researching hot" in

college students[J]. Journal of Taiyuan Urban Vocational and Technical College, 2019(06): 94-97.

[4] Shen Sudan, Zhang Huixin, Liu Xinpeng, Tang Yuqi. Research on the Problems and Strategies of College Students' Textual Research[J]. Modern Business & Trade, 2019, 40(16): 78-79.