

# Data Crawler Data Collection and QoS Optimization Path of Smart Course Online + Offline Evaluation Platform

Xi Chen

Chongqing College of Humanities, Science and Technology  
Chongqing, Hechuan, China,401524

---

**Abstract:**Based on the development background of "Internet +", this paper introduces the advantages of the application of the smart teaching mode, analyzes the problems existing in the smart teaching of vocational education courses, and explores the countermeasures for the application of the smart teaching mode in vocational education. The influence of internal factors such as educational emotion and information teaching technology and external factors such as schools and social environment, web page analysis technology, database technology, Web technology and GIS technology for extraction, storage, and spatial analysis of real estate data, "smart learning platform", etc. The core link, the routing strategy is mainly implemented by the QoS routing control framework, and the physical topology discovery module in the framework, in order to analyze the internal elements of the teaching paradigm change from the perspective of technical support.

**Keywords:** Data Crawler, Data Collection, QoS Optimization Path, Smart Course, Offline Evaluation Platform

---

## 1. INTRODUCTION

As far as vocational curriculum teaching is concerned, the application of smart teaching mode in its teaching can create a new teaching environment for vocational education, break through the limitations [1] of traditional curriculum teaching mode, and promote curriculum teaching innovation. With the rapid development of cloud computing, virtual reality (VR), holographic projection and other technologies, various modern teaching methods [2] continue to flood into the classroom, and traditional teaching is challenged. Knowledge information is freely shared through various media, making mobile learning and SNS technology [3] the mainstream of teaching development in the new century. In 1999, S.Chakrabarti first proposed the concept of "topic web crawler" at the World Wide Web Conference [4].

With the rapid development of Internet technology, the network carries a large amount of information, and how to effectively extract and utilize this information has become a huge challenge [5]. In many traditional search engines such as Google Yahoo Alta Vista and so on. Subsequently, the web crawler technology was quickly applied in various fields: Gan Guohua et al. adopted the search strategy [6] based on the improved Context Graphs method and the target page classifier method based on the support vector machine SVM for the university field to obtain useful resources [7]. The Ten-Year Development Plan for Educational Informatization (2011-2020) clearly points out the tasks and goals of improving teachers' application of information technology, building a professional technical support team [8], improving educational informatization leadership, and optimizing the information technology talent training system [9]. The Outline of the National Medium- and long-term educational reform and development plan [10].

At present, the scale of the Internet is constantly expanding, and various applications are emerging one after another. Especially in recent years [11], with the continuous maturity and development of mobile Internet technology, there has been an "explosive" increase in network users [12]. According to the 37th release by China Internet Information Center (CNNIC), the research hotspots for data center network traffic

mainly focus on network structure and routing strategy. The root node of the traditional [13] hierarchical tree network structure often becomes the bottleneck of network performance, and the number of servers is limited by the hierarchical structure [14]. The new network structure can solve this problem well. Most data center routing strategies focus on improving or redesigning routing algorithms. The total number of websites with ".CN" domain names on the Chinese Internet has exceeded 20.85 million, a year-on-year increase of 3% [15]. Faced with such a huge number of web pages, it is unrealistic to rely on the single-machine version of the web crawler to obtain enough information, even with high-performance, high-bandwidth server support. Teachers and students are not independent [16], and they are not independent. Two-way interaction can be realized. Teachers can formulate corresponding questions and answers, and students can answer. Zhou Xiaoli [17] realized a public opinion monitoring system based on general web crawler and Lucene index database. Hao Hu [18] and others used the crawler based on the Heritrix architecture to collect netizens' attitudes towards the Huangyan Island incident. Students can leave a message or ask questions online to teachers in a timely manner if they do not understand [19]. Teachers can respond in real time, and the interaction efficiency is very high. Applying the smart teaching model to vocational education teaching has many advantages [20]. When conducting network retrieval, what we hope is the largest possible network coverage, but due to the different resource allocation between the limited search engine server resources and the unlimited network data resources [21], the contradiction between them is further deepened. Information technology has a revolutionary impact on the development of education" is a major historical proposition. In the new era of comprehensive integration of technology and education [22], the construction of an innovative country urgently needs matching knowledge systems and talent training methods. Many countries in the world and the region has taken smart education as a major strategy for its future education development [23].

In the face of massive Internet data, many individuals and companies hope to obtain valuable data information or the rules contained in it. However [24], to achieve this goal,

collecting sufficiently accurate and effective data is the most basic part. The method of combining the ECMP algorithm with weights and the establishment of multi-constraint QoS network model is proposed to realize the routing strategy of the routing engine module.

## 2. THE PROPOSED METHODOLOGY

### 2.1 The Smart Course Online + Offline Evaluation Platform

At present, in addition to smart mobile terminal devices, such as smart phones, tablets, and computers, the smart teaching modes commonly used in vocational education and teaching also include smart teaching platform applications, smart teaching models, and integrated applications of courses. Smart teaching is a ubiquitous, perceptual, integrated, and intelligent new educational ecosystem built on the basis of new-generation information technologies such as mobile Internet, Internet of Things, and big data. It integrates digital technologies such as augmented reality and cloud computing, and concentrates the core of virtual reality and visualization technologies. The first is that the amount of data is too low; the second is inefficiency; the accuracy is also affected by the appraiser's personal subjective factors.

Using GIS theme crawler technology to automatically collect Web real estate transaction data and Web real estate geospatial data, and use the collected data to design an online real estate appraisal system to realize automatic real estate evaluation. It contains full text search and web crawler. The part of the web crawler adopts a plug-in-based approach to handle various types of sites, and it can also combine with doop to build a large-scale distributed crawler system. At the same time, due to the scale of modern data centers usually reaching tens of thousands or even millions of servers (as of 2006, Google has more than 450,000 servers in its more than 30 data centers. Web crawlers are often used to crawl the Internet. Of all the resources on the Internet, the most common application is the search engine. The number of web pages crawled by this type of web crawler is generally in units of hundreds of millions, and it needs to repeatedly crawl some frequently updated web pages. Crawl speed and storage capacity are extremely demanding.

With the application of smart teaching mode, efficient resource sharing can be achieved when connected to the network. The application of the smart teaching model can effectively break through the limitations of time and space, allowing teachers and students to achieve remote interaction and shorten the distance between teachers and students, even in special circumstances

### 2.2 The Data Crawler Data Collection

Property geographic information data includes property location information and map basemaps. Location information collection is collected by using the huge data source of the Web, which is more efficient than collecting in the wild. In this paper, the dynamic web page information acquisition technology is used to simulate the user click query process, and the coordinate data of the corresponding point is returned. The crawler system can automatically download relevant web pages, and formulate the targets to be crawled according to the corresponding information and the instructions issued by people. It can selectively access web pages and related links on the Internet, and then decrypt the required information data. This web crawler adopts a neutralized topology, in

which the task distribution node is the master node and the other nodes that accept tasks are the work nodes.

Through the above analysis, it is concluded that the Web real estate data source has the following characteristics: First, there are a large number of noises such as navigation bars, advertising links, copyright notices, etc. in the webpage; Second, the data carrier is divided into static pages and dynamic pages, which need to be combined with different page extraction technologies. All web pages crawled by web crawlers will be stored, filtered, analyzed, classified and indexed by the system. The biggest difference between the topic web crawler and the general web crawler is that the topic web crawler analyzes the content of the target web page during the crawling process and judges its relevance to a specific topic.

### 2.3 The Qos Optimization Path of Smart Course Online + Offline Evaluation Platform

Integrated Services (IntServ, Integrated Service), differentiated services (DiffServ, Differential Service) and traffic engineering (TE, Traffic Engineering). The IntServ model introduces the Resource Reservation Protocol (RSVP), which enables network applications to achieve end-to-end QoS guarantee, but its scalability is poor. On the basis of the original directional crawling, the whole network crawling is added, and then the Topic recognition technology, web page slicing technology and web page weight analysis technology, so as to achieve the function of automatically tracking links and extracting text by crawlers. Compared with traditional classroom teaching, online learning has the characteristics of autonomy and collaboration. Through certain restraint and supervision mechanisms, it can not only collect information such as students' learning behaviors, but also provide data support for subsequent evaluation mechanisms.

Then add the support of the robots protocol to the crawler to make our crawling legal. Under the background of "Internet +", vocational education teachers need a certain period of time to master information technology, and they must maintain a good attitude and gradually master the skills and knowledge. During the learning period, it is inefficient for teachers to explore behind closed doors alone, and they may also take detours. In order to improve the efficiency of smart education in vocational colleges, it is necessary to fundamentally increase the importance attached to smart education by the majority of vocational teachers, to actively carry out smart education teaching guidance work conferences, and to publicize the importance of smart education. The central engine, as the core brain of Scrapy, It is responsible for managing all data flows in the system. The input and output of each module need to go through the central engine, so this component occupies a very important position in the Scrapy framework, once the central engine fails.

In the formula, IB is the popularity of the link, which is determined by the quantity and quality of backlinks; IL is the importance of the link, which is a function of the URL string; X and Y are the weight coefficients of IB and IL. The principle of the template processor is that the template is composed of some modules, each of which handles different tasks.

### 3. CONCLUSIONS

Vocational educators should adhere to the principles of "going fast and thinking carefully", "things change people's stability", and "secret education in Tibetan education", with the development of learners' subjectivity as the core, generative learning activities as the main body, and ubiquitous learning resources as the basis. The automatic analysis and processing of the seed list has been realized, and the attempt of intelligent collection of the website has been carried out, and certain achievements have been obtained. At the same time, the traditional crawler does not support the collection.

### 4. REFERENCES

- [1] Peng Shanshan. The teaching path of smart classroom under the "Internet +" mode [J]. *New Curriculum Research*, 2020(21):2.
- [2] Kong Lingbiao, Cui Jie, Yang Ming, et al. Intelligent QoS routing optimization method based on deep reinforcement learning in SDN environment, System:.
- [3] Cao Yingzhuo. Dynamic Evaluation Mechanism and Protocol of Overlapping Network Path QoS Capability [D]. Northeastern University, 2013.
- [4] Han Lifeng. Service Quality Optimization Method and System, Network Element: CN102740379A[P]. 2012.
- [5] Wang Yue. Design and implementation of service quality control system based on software-defined network [D]. Beijing Jiaotong University, 2017.
- [6] Zhao Meili, Tang Jing. Exploration on the Smart Learning Path of University Computer Basic Courses from the Perspective of "5G+AI" [J]. *Computer Education*, 2022(1):5.
- [7] Liu Yangxi, Li Jiabin, Yuan Mengdi. The optimization path of smart party building in colleges and universities in the era of "Internet +" [J]. *Journal of Hubei University of Economics: Humanities and Social Sciences*, 2021, 18(4):4.
- [8] He Wenlian, Zhuo Wenxia, Lin Luchun, et al. Analysis of the quality improvement path of "Internet +" smart elderly care services in Guangdong Province [J]. 2020.
- [9] Lin Baoshu, Paul B S. In the 4G broadband mobile system environment, a discussion on the mobile augmented reality technology and application using smart devices as a platform - the overall project and sub-project three researched and developed in the 4G BMS environment QoS and Network Performance Evaluation/Analysis of SD-MAR Application. 2012.
- [10] Zhang Baoji. Research on Integrated Audio and Video Service Quality and Intellectual Property Guarantee under Wireless Network (II); Integrated QoS and IPR Guarantees for Audio/Video Applications over Wireless Networks (II) [J]. Executive Yuan National Scientific Committee, 2010.
- [11] Cao Yingzhuo. Dynamic Evaluation Mechanism and Protocol of QoS Capability of Overlapping Network Paths [D]. Northeastern University, 2014.
- [12] Hu Jie. Network teaching platform: connotation, characteristics and innovative development path [J]. *Decision and Information*, 2018(6):10.
- [13] Wang Jinlan. Research on wavelength path selection technology based on quality of service in SUPANET [D]. Southwest Jiaotong University.
- [14] Wang Xianghui. Semantic web service composition evaluation and optimization under QoS constraints [J]. *Journal of Tianjin University: Natural Science and Engineering Technology Edition*, 2015, 48(2):13.
- [15] Liu Jie. Taking science and technology and innovation as endogenous driving forces to optimize the development path of smart education: Smart education in Haidian District, Beijing is moving towards the 2.0 era [J]. 2022(9).
- [16] Qi Le. Research on composition and optimization of logistics cloud services based on QoS [D]. Hefei University of Technology, 2014.
- [17] Chen Jianyong, Hu Cunying. System and method for network optimization control based on security and service quality: CN101854337B[P]. 2012.
- [18] Wang Jianxin, Wang Xinhui, Peng Gegang. An Efficient Multi-Restricted Optimal Path Selection Algorithm in QoS Routing [J]. *Computer Science*, 2002, 29(12):4.
- [19] Sun Jing. Strategies for optimization and improvement of water supply service quality under the background of smart city [J]. *Economic Management Digest*, 2020(18):2.
- [20] Zhang Baohou. Research on NB-IoT network capacity estimation and optimization technology [D]. Beijing University of Posts and Telecommunications, 2019.
- [21] Chen Menglin. Research on service quality optimization of offline smart outlets of Z bank C branch.
- [22] Wang Yafei. Research on the application of QoS mechanism in communication platform in smart scenic spots [D]. Beijing Institute of Technology.
- [23] Xi Wenwen. Optimization of educational management paths in higher vocational colleges from the perspective of smart education [J]. 2020.
- [24] Yuan Yijie. "Internet +" smart education information platform optimization design research [D]. Hebei University of Science and Technology.