Journal of Contemporary Educational Research, 2025, Volume 9, Issue 7
http://ojs.bbwpublisher.com/index.php/JCER
ISSN Online: 2208-8474
ISSN Print: 2208-8466

Teaching Reform and Practice of the "Data Collection and Web Crawler" Course Based on the Blended Teaching Mode

Simin Wu*

Department of Computer Science, Guangdong University of Science and Technology, Dongguan, China

**Author to whom correspondence should be addressed.*

Abstract: The data collection and web crawling course has a lot of theoretical knowledge and strong practicality. Traditional teaching methods are no longer sufficient to meet teaching needs. Based on the characteristics of the course, this article constructs a mixed teaching environment based on "Learning Pass + Hongya Platform + Offline Course," integrates teaching resource libraries and ideological and political cases, and develops a suitable evaluation system to cultivate students' innovative and critical thinking abilities, stimulate their learning initiative, improve their teamwork ability, and enhance their professional level and data literacy.

Keywords: Blended learning mode; Crawler; Course teaching reform

Online publication: July 31, 2025

1. Introduction

With the rise of more and more online teaching and experimental platforms, as well as the widespread popularity of online teaching models during the epidemic period, blended learning models are widely used in university course teaching, implementing the "student-centered" approach in practice, improving the drawbacks of traditional teaching models, deepening teaching reforms, and greatly improving the overall effectiveness of computer course teaching.

Data Collection and Web Crawler is a professional course for undergraduate majors in computer science and technology, data science and big data technology, etc. It involves crawling static and dynamic web page source code, data parsing, data storage, and Scrapy framework crawling. Not only do students need to master solid knowledge of web crawling, but they also need to master relevant web crawling tools and write relevant code. Through the study of this course, students will be able to master the basic knowledge of Python crawlers, independently develop Python crawler project programs using frameworks, independently complete data crawling work, and provide corresponding solutions for obtaining complex data. This will lay a good foundation for future professional course learning and data-related work.

Blended learning is a teaching model that combines online and offline learning, including online learning and traditional classroom teaching [1]. This teaching mode requires students to first study videos or designated materials uploaded by teachers online, preview the teaching content in advance, and then discuss learning with teachers in class to solve problems in a timely manner, or combine theoretical knowledge previewed in class to better carry out classroom project cases and overall improve students' learning effectiveness. The role of a teacher is to guide, inspire, and manage students, stimulate their enthusiasm, initiative, and creativity in learning courses, encourage their learning, thinking, and entry into deeper levels of learning [2].

Chaoxing Learning Platform integrates rich educational resources, which are not only suitable for students and teachers but also provide educators with abundant teaching resources and interactive tools. Users can efficiently learn and manage anytime, anywhere. With the help of Chaoxing Learning Platform, teachers can quickly create online courses, conduct teaching interactions, evaluate learning, grade assignments, etc. Therefore, the blended teaching practice based on Chaoxing Learning Platform has attracted much attention from the education industry. Many teachers and students in universities have almost downloaded the online teaching platform Xuetangtong and completed various online teaching activities on it.

At present, online experimental platforms are widely used, such as 360AI development platform, Hongya platform, Touge experimental platform, and experimental platforms independently developed by major universities. The Hongya platform used in this article is an online experimental platform related to the big data major at Guangdong University of Science and Technology. It has rich online resources and experimental platforms for big data-related courses, detailed theoretical knowledge, and experimental operation documents, allowing students to complete web crawling projects directly on the platform's virtual machine without setting up an environment. The courses related to this course on the Hongya platform include "Python Fundamentals and Applications," "Python Crawler," "Introduction to Big Data," "MySQL Database," "MongoDB Database," etc. These related courses will help students learn web crawling courses more systematically.

2. Characteristics and current status of the course

2.1. Characteristics of the current course

The course on data collection and web crawling has a wide range of content, a strong theoretical basis, and a complex knowledge system. It involves prerequisite courses such as introduction to big data, program design, relational databases, non-relational databases, and web design. It is highly comprehensive and requires a large amount of teaching resources. The theoretical and practical aspects of the teaching process need to be closely integrated, but there are some problems in traditional teaching.

2.1.1. Outdated teaching methods and content

In the teaching process, teachers tend to focus more on the completeness, systematicity, rigor, and rationality of the knowledge system in web crawling courses, but overlook the "student-centered" and "rapidly developing demand for data talents." In traditional teaching, the correspondence between the content (theory and practice) taught by teachers and the abilities that students should possess is not clear enough, and the integration of ideological and political modules is not appropriate, failing to timely establish correct data literacy for all students. Due to the updates and iterations of web crawling technology and the continuous changes in website page structure, some teaching video resources, textbooks, materials, and teaching tool versions have been replaced, resulting in the inability to crawl corresponding website data. Additionally, the lack of timely updates to web crawling teaching resources can also affect the overall learning effectiveness.

2.1.2. Lack of rationality in practical teaching

At present, most universities in China have offered courses on data collection and web crawling, but there is a lack of systematic and diversified tools and methods, with many problems in the implementation stage of teaching practice [3]. The lack of a comprehensive connection between theory and practice in course teaching, as well as insufficient cross-integration between web crawling courses and other courses, can result in students being unable to fully master web crawling technology. Online teaching platforms such as Xuetangtong lack unified online and offline resources and corresponding reference materials. The simple hybrid mode of offline courses and Xuetangtong makes it difficult for teachers to understand the actual practical teaching effects of all students and lacks rationality.

2.1.3. Single course evaluation systems

Currently, some universities still use the final exam as the sole criterion for assessing web crawling courses. Simply analyzing the exam papers or a web crawling project cannot comprehensively evaluate students, and it also limits their enthusiasm for learning the course, resulting in students only completing their work and reviewing exam content. This evaluation system cannot intuitively reflect students' hands-on ability, cannot comprehensively reflect students' learning outcomes, and cannot improve students' data processing skills. As data collection and web crawling is a highly practical course, the evaluation system needs to not only consider students' understanding of basic knowledge, but also assess their ability to design crawling rules and solve code bugs in actual projects.

2.2. Current learning characteristics of undergraduate students

In the early stage of COVID-19, the online teaching of this course has accumulated more teaching resources, teaching design, and teaching videos, which are more in line with the development of the current new era. Students have also adapted to the use of electronic notes, cloud storage, online platforms, and the sharing of teaching resources to a certain extent [4].

The current situation of students is that there are undergraduate students without a foundation in web crawling, students who have advanced to undergraduate universities, and students who have some foundation in web crawling. The first type of students needs to include basic web crawling concepts and webpage structures in their teaching plans. The second type of students already have a certain understanding of the basic concepts and techniques of web crawling courses and introductory courses, and can directly participate in the web crawling project teaching. Secondly, students vary from person to person, with different learning situations, knowledge reserves, thinking styles, logical abilities, and hands-on abilities, which makes the course implementation somewhat difficult. Teachers generally design teaching plans, activities, and modes based on the general level of students, which may result in some students' learning progress being slower than others, thereby affecting the overall teaching effectiveness of the course [5].

3. The construction process of blended learning between online and offline teaching

This article has reformed the existing teaching mode, combined with the existing resources of the school, integrated multiple teaching modes for teaching work, and constructed a mixed teaching mode of "learning general education–Hongya experimental platform–offline classroom case teaching." It has been applied to subsequent classes, continuously improved and summarized, in order to better improve the teaching quality of the course.

3.1. Integrating online and offline teaching resources

The integration of teaching resources, guided by learning outcomes in curriculum construction, optimizes and adjusts teaching content, teaching focus, and teaching difficulties, establishes a curriculum teaching outline based on blended learning mode, carries out a series of teaching designs and practical training guides, organically integrates web crawling projects into classroom teaching, connects theoretical teaching and practical activities, and cultivates students' engineering abilities. Introduce precise web crawling ideological and political project cases in classroom teaching design, including introducing relevant school websites, Baidu websites, Learning Platform, Douban platform, etc.; Introduce the history and importance of related web crawling technologies, and stimulate students' innovative abilities; Guide students to establish correct values by introducing characters related to classroom content; By introducing associations between this course and other courses, such as associating knowledge points with relational and non-relational databases when storing data after web crawling. Establish a data collection and web crawling course on the Learning Platform combined with the Hongya Platform, providing students with rich teaching resources such as PPTs, instructional videos, teaching outlines, and practical training guides. At the same time, carry out activities such as homework, Q&A, discussions, and tests on the Learning Platform to better achieve blended online and offline teaching of Learning Platform + Hongya Platform + classroom.

3.2. Building a blended online and offline teaching model of Xuetong + Hongya platform + classroom

The blended learning mode of this article is shown in **Figure 1**.

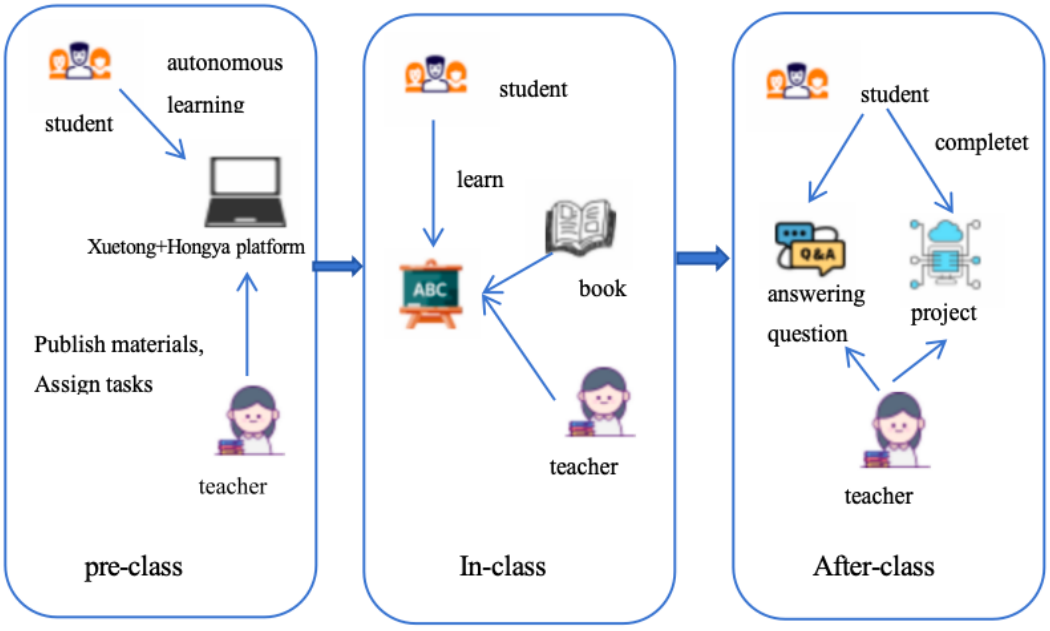

Figure 1. The teaching model of blended learning

Before teaching, learn the course outline, teaching syllabus, and teaching plan for showcasing data collection and web crawling to students, including the abilities required for the course, the key and difficult points of each chapter, crawling cases in each chapter, teaching methods, expected goals, assessment methods, time allocation, etc. Before class, the teacher assigns learning tasks on the Hongya platform for students to preview theoretical knowledge in advance, complete theoretical knowledge quizzes, enhance their

understanding of theoretical knowledge, and complete the crawling project for current knowledge points according to the experimental operation steps prepared on the platform. This facilitates students to preview the upcoming crawling knowledge points in advance and practice familiarizing themselves with the upcoming crawling project. The two online platforms, Xuetong and Hongya, before teaching, help students with different backgrounds to master knowledge points and lay a good foundation for offline classroom teaching in the later stage.

In teaching, combined with a blended learning mode, teaching methods such as problem-based learning, case-based teaching, project-based teaching, and task-driven teaching are adopted for different teaching contents, and the classroom atmosphere is enlivened. The teacher demonstrated the current classroom case to attract students' interest in learning, solved the problems encountered in the Hongya Platform crawler project, and independently attempted to crawl classroom project data. Due to the pre-teaching tasks assigned, teachers can group students with different student backgrounds in class and have representatives from each group demonstrate their work, allowing students to discuss and learn from each other, thereby improving their teamwork and project management skills. Teachers guide students offline in class to complete specific web crawling projects, promptly solve problems encountered by students, and have a good grasp of the required web crawling techniques. In addition, teachers can assign classroom practice tasks that allow students to try to think about crawler rules from different dimensions, triggering deep learning among students and promoting their classroom participation, thus achieving a mixed mode of online and offline teaching.

After the lecture, after pre reading and offline teaching activities, teachers can view the learning situation of all students and understand their existing problems through the management pages of the Learning Platform and Hongya Platform. For students with good and moderate learning progress, they can be assigned more intensive additional homework to improve their crawling ability. For students with difficult learning progress, guidance can be given. For students with a less positive learning attitude, timely warnings can be given. Teachers can assign crawling homework that extends classroom knowledge according to students' learning situation based on easy and difficult problems. Secondly, teachers should establish learning and communication groups through Study Connect and DingTalk to promptly address students' problems and obtain feedback on the current course, which enables teachers to continuously summarize and improve the teaching of the course. Teachers can also stimulate students' innovation ability and ability to solve practical crawling problems by introducing some scientific research projects on crawling tasks, thus realizing the blended online and offline teaching mode of "Learning Pass + Hongya Platform + Offline Classroom" for this course.

3.3. Construction of a course evaluation system

A complete evaluation system not only objectively and truthfully reflects the effectiveness of learners' personal learning of web crawling courses, but also identifies their own problems and conducts targeted active learning, ultimately leading to the achievement of teaching objectives. Teachers can better evaluate students' learning progress through the backend management pages of online platforms such as Hongya Platform and Learning Platform. This project is based on a blended learning model combined with practical implementation to develop an evaluation system for this course. The evaluation indicators for the course include theoretical knowledge assessment, practical ability assessment, innovation ability assessment, process assessment, and other aspects. According to the actual talent cultivation plan, propose an evaluation strategy of 20% for the learning situation on the Hongya platform, 20% for the learning situation on Xuetong, 20% for offline classroom performance, and 40% for final assignments, in order to comprehensively evaluate students.

3.4. Summary of examples of teaching cases

Existing teaching resources: Xuetangtong has already released teaching progress, syllabus, lesson plans, teaching experiment reports, teaching courseware, and necessary software to students. Xuetangtong has uploaded operation videos for most tasks. All students on the Hongya platform have accounts and can access data collection and web crawling-related content on the platform. The teaching knowledge point selected as an example in this article is the regular expression parsing of data in the course of data collection and web crawling. Before class, the teacher posted a simple questionnaire survey on the Learning Platform and a regular expression task on the Hongya platform. There are 5 multiple-choice questions on regular expressions and 3 cases of regular expression processing string text on the Hongya platform. The teacher can check if students have already learned about regular expression knowledge in other courses. Enable students who have already learned to briefly review previous knowledge points, while students who have not yet learned can use the Hongya platform to learn and then answer questions, allowing students to understand in advance the knowledge points that will be learned in the next class. Teachers review students' questionnaire responses and learning progress on the Hongya platform, and choose teaching methods reasonably. Because regular expression parsing of page data is a highly practical knowledge point, and since the homework of crawling the source code of the school's official website has already been assigned, teachers can use regular expression parsing of data in webpage source code as a project case for offline teaching in the Learning Platform, such as parsing strings, partial hyperlinks, phone numbers, email addresses, etc. of the "School of Computer Science." During the offline teaching process, teachers explain multiple-choice questions on the Hongya platform and provide simple cases of regular expression parsing data. Students are grouped to complete and demonstrate a web crawling project on-site. Teachers patrol and solve students' problems before and during class, score and evaluate their presentations, and automatically record them on the Learning Platform as evidence for the later evaluation system. Teachers summarize classroom content, clarify common problems among students, and assign homework projects after class. After class, teachers and students maintain communication and learn together. Teachers can continue to observe students' learning progress in the background, and students can also seek advice from teachers in a timely manner.

4. Conclusion

This article is guided by a blended learning teaching model, with learning as the center, allowing students to actively participate in the learning of the course "Data Collection and Web Crawling," enabling them to complete various learning tasks and see their own achievements, thus becoming active learners; Teachers can use the platform to focus on students' learning outcomes, enhance their subjective initiative, self-learning ability, and proactive acquisition of new knowledge. This article integrates teaching resources and course ideological and political cases, using a blended learning model based on "online learning platform + online Hongya platform + offline classroom teaching" before and after teaching, to construct a course evaluation system based on the blended learning model, enabling students to master basic web crawling skills and participate in scientific research project development, thereby cultivating students' scientific research ability and improving their ability to collaborate with groups.

Funding

This work is supported by the Quality Engineering Project of Guangdong University of Science and Technology

under Grant GKZLGC2024160.

Disclosure statement

The author declares no conflict of interest.

References

[1] Bian Q, 2019, Exploration of Teaching "Data Collection of Automotive Engine Bracket" Based on Blended Teaching Mode. China Equipment Engineering, (24): 211–213.

[2] Liu M, Zheng R, Wang Y, 2023, Exploration of the Practice of Blended Teaching Mode in Computer Network Courses. Computer Knowledge and Technology, 19(17): 89–91.

[3] Liu J, 2023, Research on the Blended Online and Offline Teaching Mode of the Course "Computer Networks." China New Communications, 25(15): 83–85.

[4] Liao B, 2024, Exploration and Practice of Blended Teaching Mode in Computer Graphics. China Educational Technology Equipment, (06): 113–115.

[5] Wu X, Lin Q, Zhou Y, 2024, Research on Blended Online and Offline Teaching of "Computer Programming" Course. Journal of Huainan Vocational and Technical College, 24(04): 79–81.