Enhancing and Analyzing Log Generation for Collaborative Problem-Solving Activities: Video Analysis and OCR Techniques

Suprabha Jadhav[1](✉), Parth Jain[2], Harjot Singh Raith[3], Kavi Arya[4], and Sridhar Iyer[1]

[1] Educational Technology, IIT Bombay, Mumbai, India
suprabhaj@iitb.ac.in
[2] Information Technology, Atharva College of Engineering, Malad, India
[3] Artificial Intelligence and Data Science, NMIMS, Navi Mumbai, Kharghar, India
[4] Computer Science and Engineering, IIT Bombay, Mumbai, India

Abstract. This study explores the use of video analysis and Optical Character Recognition (OCR) to generate accurate logs for tracking user search behaviors for external resources during collaborative problem-solving activities on the RoboReady educational platform. The research included 15 teams of engineering students who interacted with a website designed around the PISA 2015 collaborative problem-solving framework. Video recordings of the user's screens were analyzed using image processing and OCR to extract URLs and timestamps of visited web pages. The study compares these logs with those collected through server-side tracking, noting the limitations of the latter in capturing external resource usage. The methodology includes frame extraction, URL isolation, and text recognition using Tesseract OCR and Google Cloud Vision API. Challenges like noisy OCR output and the necessity for manual verification are addressed. The generated logs provide insights into user's navigation and resource usage. However, limitations such as video quality dependency, scalability issues, and the need for manual intervention are also discussed. This research enhances the understanding of user's behavior in online learning environments and offers methodological insights for studying digital learning interactions.

Keywords: Collaborative Problem-Solving · Educational Data Mining · Optical Character Recognition · Video Analysis · User Behavior Tracking

1 Introduction

User behavior logs provide crucial insights into educational platform usage, revealing study habits, content popularity, and engagement patterns. These insights can inform improvements in site structure, content delivery, and student support.

Pattern recognition advancements, particularly in Optical Character Recognition (OCR), have significantly improved document processing efficiency across various industries. With accuracy rates exceeding 99% for printed text, OCR technology has become widely adopted [5].

A. Karkare et al. (Eds.): COMPUTE 2024, CCIS 2400, pp. 135–143, 2025.
https://doi.org/10.1007/978-3-031-84391-4_11

Educational websites have evolved into interactive learning tools, offering personalized and flexible opportunities. Understanding user interactions with videos and content is crucial for predicting performance and improving educational outcomes [1]. Advanced analytics of video and text resources can inform content development, platform optimization, and teaching methods. The growing popularity of free and paid educational platforms provides opportunities to identify best practices for enhancing online learning experiences through comparative analysis [2]. This research aims to understand the role of educational websites in modern education and how data analytics can improve their effectiveness. It continues a previous study [9] that focused on fostering teams' problem-solving skills, with the current study emphasizing collaborative problem-solving skills.

2 Literature Review

In [1], student interactions with video sequences were analyzed from 66 students in a C++ Moodle course, examining 7,423 clicks. K-Nearest Neighbors (KNN) and Multilayer Perceptron (MLP) algorithms were used, achieving 65.07% and 61.13% accuracy, respectively, showing that video interaction patterns can predict student performance. In [2], methods for improving educational video navigation were introduced, including a "Rollercoaster" timeline for visualizing navigation, an enhanced in-video search feature, and a video summarization method highlighting frequently viewed frames.

In [3], web mining techniques were applied to analyze web usage data from www.davkota.org, identifying key metrics like 23,669 hits and 935 unique IP addresses. [4] proposed a privacy-preserving method for mobile screen recordings with on-device event detection, preferred by users over server-based approaches. Meanwhile, [5] highlighted the importance of preprocessing techniques such as noise removal and skew correction in character recognition, improving real-time data extraction.

Methods for extracting visited domains from screen recordings were discussed in [6], utilizing techniques like frame extraction, template matching, and OCR. [7] introduced Web Historian, a Chrome extension for analyzing browsing history with visualization and deletion capabilities. Lastly, [8] explored using perceptual hashes, template matching, and OCR to detect UI events in screen recordings, though OCR accuracy posed challenges during transitions.

3 Methodology

A study was conducted using the 'RoboReady', an educational platform designed to fostered students Collaborative Problem-Solving (CPS) skills. The study had following research questions:

RQ1. How effective is the use of video analysis and OCR techniques in generating accurate log data for tracking undergraduate students' searches for external scaffolds during collaborative problem-solving activities?

RQ2. How can extracted log data be utilized to enhance the understanding of resource utilization during collaborative problem-solving activities?

3.1 Participants

The study involved a total of 15 teams comprising undergraduate students from various departments and years of study. Specifically, participants included students from the Electronics & Computer Engineering department in their second year, Computer and Computer Engineering departments across second and third years, as well as from MBA Tech (B Tech in Computer Engineering + MBA). Additionally, students from Mechanical, Electronics and Telecommunication (EXTC), and Computer Science departments, ranging from first year to third year, also participated. The diversity in departments and years of study provided a broad perspective for the research conducted. Each team comprised of 2–4 students.

3.2 Activity

Participants were required to navigate through a website structured around the PISA 2015 collaborative problem-solving framework [10]. Each stage of this framework included specific activities and quiz questions. To support these tasks, each stage featured a resources page containing various scaffolds, such as conceptual videos, worked examples, and relevant website links, supplemented by additional resources. Teams were tasked with completing activities, including quizzes using the provided scaffolds. With individual student consent, all interactions on the website were logged on the server, and their screens were recorded. This enabled the creation of user log files for analyzing [10] behavioral patterns during the activity.

The video recordings captured both the students' browser windows and live camera feeds. These recordings were analysed using Image Processing tools and Optical Character Recognition (OCR) to extract the URLs of the web pages visited and the corresponding timestamps whenever the screen content changed. This process provided a comprehensive log of teams' behaviour throughout the activity.

3.3 Log Data Collection

Structure of the Log Data. The log data structure, as illustrated in Fig. 1, consists of three main components: user_id, which serves as a unique identifier for each team and acts as the primary key; Page_description, representing the name of the website subsection being accessed; and Timestamp, indicating the date and time of access for a specific page section.

User_id	Page_description	Timestamp
714	Stage 3: Activity	02-03-2024 01:55
723	RoboReady	02-03-2024 02:11
724	RoboReady Website	02-03-2024 02:12
725	Stage 2: Resources	02-03-2024 02:14
727	Stage 4: Activity	02-03-2024 02:14

Fig. 1. Sample server log data

Data Cleaning. To enhance data quality, three main steps were implemented: sorting the data by user_id to facilitate systematic analysis, renaming extracted URLs to their corresponding page titles, and calculating dwell_time by determining the differences between consecutive timestamps.

Limitations. The server-based log collection process has several limitations that affect analysis accuracy: incomplete data collection, as only interactions within the website were captured, the use of external resources, such as Google, YouTube, and AI tools, which were not recorded by the server; and the impact on data accuracy, as the use of external resources introduces significant variance in the calculated engagement time, leading to an underestimation of the actual time spent on activities.

3.4 Generating User Activity Logs from Video Recordings

To overcome server-based log collection constraints, we analyzed video recordings to obtain log data, enabling understanding of external scaffold usage. Figure 2 outlines the workflow: frame extraction, URL extraction, OCR processing, and data cleaning.

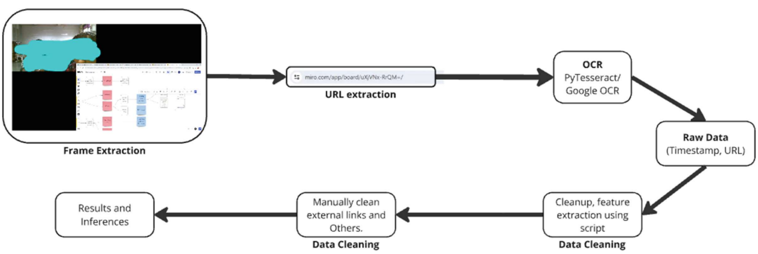

Fig. 2. Log generation process

Data Collection Process. Log file creation from video recordings employed Python libraries, including OpenCV (cv2), NumPy, and Tesseract (Pytesseract), or alternatively, Google Cloud Vision API for OCR. Frames were extracted every second, and those with a mean squared error (MSE) exceeding a threshold of 20 compared to the previous frame underwent further processing, with this threshold determined experimentally. URLs were isolated by identifying the URL bar coordinates using either template matching [6] or manual extraction if the position remained consistent. Preprocessing steps [5], such as grayscale conversion, upscaling, dilation, and sharpening, were applied to enhance OCR accuracy. For text extraction, Tesseract OCR (configured with --psm 6) performed well with high-quality images, while Google Cloud Vision API offered superior accuracy for lower-quality frames, albeit at a higher cost.

Data Cleaning. The procedures generated CSV log files with timestamp and URL columns for each video, though URL accuracy was often compromised due to poor video quality. To address this, we implemented several data cleaning steps. We first discarded redundant data by eliminating consecutive rows with identical URLs and empty URL rows. Next, we matched extracted URLs against a compiled list of known websites and scaffold URLs using the Levenshtein distance metric [11], creating a 'cleaned_url' column. We then categorized known URLs into subtopic lists and matched cleaned URLs against these, creating a 'sub_topic' column with detailed resource information. For external searches, we used the urllib library to parse noisy URLs, extract query parameters, and identify user search inputs. We cleaned these inputs and employed the Llama3-70b LLM to refine and infer likely search terms. Finally, for instances where OCR failed to provide clear URLs, we performed manual verification by reviewing video timestamps and populating remaining fields.

Figure 3 illustrates the keyword extraction process and compares OCR technique accuracy.

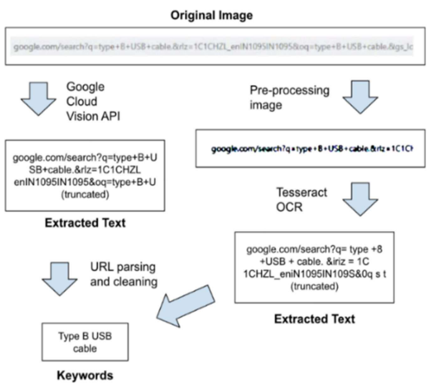

Fig. 3. Example process of keyword extraction

4 Results and Discussions

4.1 Addressing RQ1

The method in this study captured both on-platform activities and external searches, offering deeper insights than server logs alone. By integrating video analysis and OCR, it generated comprehensive logs reflecting a full range of student interactions, including external resources. This provided a more accurate representation of student behavior during collaborative problem-solving on the RoboReady platform.

As shown in Fig. 4, the integration of video analysis and OCR techniques captured detailed records of student activities, including searches for external resources. The log data identifies accessed resources and provides contextual information regarding the timing and nature of these interactions.

Each entry in the log encompasses the key components detailed in Table 1.

user_id	time(in sec)	page	class	cleaned_page	sub_topic	keywords	cleaned_k	page_desc	timestamp	dwell_time
108	4966	robo.personalt	robo	https://robo.personaltutoring.in/resources				Stage 1: Activity	0 days 01:22:46	5
108	4971	docs.google.co	Stage 1: Ac	https://docs.google.com/forms/d/e/1FAIpQLSfBy-hhv	Stage 1: Forms			Stage 1: Activity	0 days 01:22:51	3
108	4974	robo.personalt	robo	https://robo.personaltutoring.in/resources				Stage 1: Activity	0 days 01:22:54	205
108	5179	drive.google.co	Stage 2: Re	https://drive.google.com/file/d/1X0j0vlH6vfGDADcfig	Stage 2: Concept Map			Stage 2: Resources	0 days 01:26:19	170
108	5349	robo.personalt	robo	https://robo.personaltutoring.in/task-stage-2				Stage 2: Resources	0 days 01:29:09	7
108	5356	images.edrawn	Stage 2: Re	https://images.edrawmind.com/templates/cell-conce	Stage 2: Concept Map			Stage 2: Resources	0 days 01:29:16	33
108	5389	robo.personalt	robo	https://robo.personaltutoring.in/task-stage-2				Stage 2: Resources	0 days 01:29:49	4
108	5393	researchgate.n	Stage 2: Re	https://www.researchgate.net/profile/Camila-Cicuto/	Stage 2: Concept Map			Stage 2: Resources	0 days 01:29:53	80

Fig. 4. Sample log data generated from the video recordings

Table 1. Structure of the log data generated using video recordings

Column name	Description
time	Timestamp in seconds
url	Raw URL, extracted using OCR
class	Rough categorization of the data
cleaned_url	Cleaned URL for known webpages
sub_topic	The description for links provided within the website
keywords	Noisy keywords extracted from the URL
cleaned_keywords	Cleaned keywords generated using an LLM
page_desc	Readable title for the known webpages
time_in_minutes	Human readable timestamps
dwell_time	The time spent on the current page

4.2 Addressing RQ2

This comprehensive log data provides valuable insights into student engagement and resource utilization, aiding in the improvement of educational practices and resource design. It supports the creation of detailed process models and dwell time analysis, offering a nuanced understanding of student activity. Process models depict the sequence and flow of actions, while dwell time graphs show the duration spent on specific tasks. Figure 5 presents a sample process model for all teams. In this process model, the numbers on the nodes indicate the frequency of visits to each page, while the numbers on the arcs represent the frequency of transitions between pages. This quantification helps identify popular pages and common navigation patterns among students.

These process models can provide several insights:

- Engagement Pattern Analysis: The models highlight common sequences followed by students, revealing preferred engagement methods. In Fig. 5, the sequential structure suggests students generally accessed pages in the intended order.
- Resource Utilization: As illustrated in Fig. 5, the 'Stage 2: Resources' page was frequently accessed alongside 'Stage 2: Activity,' with longer dwell times indicating that students found these resources relevant. However, further analysis is needed to assess their effectiveness in supporting learning outcomes.

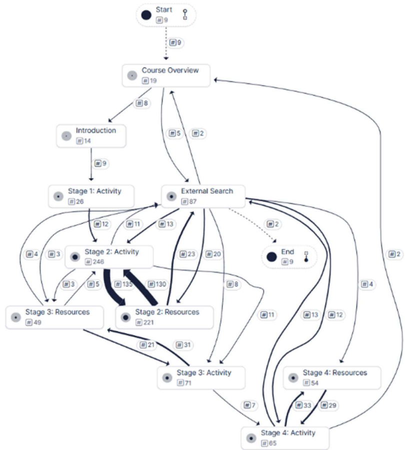

Fig. 5. Combined process model of all the teams

- External Resource Integration: Tracking external searches offers insights into additional information students seek, which can inform improvements to existing resources or the integration of commonly searched external content into the platform.

4.3 Educational Implications

The integration of video analysis and OCR for log generation has significant implications for computing education, particularly in enhancing learning analytics and supporting more effective problem-solving approaches. For example, analyzing search behaviors and resource use during debugging or algorithm design tasks can provide valuable insights into students' learning processes and strategies.

In collaborative learning environments, such as team-based coding or software development projects, this approach can offer deeper understanding of group dynamics and how students engage with external resources like coding forums or documentation repositories. Educators can identify patterns that suggest over-reliance on external tools, signalling potential gaps in foundational knowledge that may need to be addressed through curriculum adjustments or targeted interventions.

Moreover, the data generated through this method can be integrated into automated feedback systems, enabling real-time guidance for students as they engage with computational tasks. Additionally, such data can inform predictive algorithms that flag students who may be at risk of underperforming, allowing for timely interventions in both in-person and online learning environments.

5 Limitations

While our method provides rich interaction data, frequency of access alone doesn't necessarily indicate engagement quality or learning outcomes. Future research could combine this log data with qualitative assessments for a more comprehensive understanding of the learning process in computing education.

Scalability is a limitation, as this data collection method is feasible only for smaller sample sizes. For larger-scale implementations, browser extensions like Web Historian could be used, though they can't provide information about behavior within webpages.

Video quality significantly affects OCR accuracy, with lower quality compromising log data. While Google Cloud Vision API improved performance, it's a paid service. Open-source alternatives like Tesseract could be considered for cost-effectiveness.

6 Conclusion and Future Work

In conclusion, the use of Optical Character Recognition (OCR) combined with process models and data analysis techniques provides a powerful approach to understanding and enhancing educational systems. Our approach provides a more nuanced understanding of resource utilization and learning patterns than traditional server logs alone.

Future research will focus on enhancing data collection methods, interaction techniques, and analysis for improved understanding and usability of digital platforms. Expanding our current project, which uses OCR methods to extract log data from video recordings, we aim to support multiple browsers and mobile devices through new extensions and applications. Ethical data collection practices will be prioritized by allowing participants to review and selectively delete sensitive data before submission, while advanced filtering algorithms will enhance data accuracy. Web usage mining techniques will be applied to analyze metrics like hits, page views, and visitor patterns, providing insights for improving website usability and design. By integrating these advanced techniques, our research will contribute to a comprehensive and ethical understanding of user behavior, promoting the development of user-centric digital platforms.

Acknowledgments. The authors express their gratitude to the participants who took part in the study and provided us with valuable feedback. This study was conducted in the e-Yantra Lab, a project funded by the Ministry of Education (MoE).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. El Aouifi, H., El Hajji, M., Es-Saady, Y., Douzi, H.: Predicting learner's performance through video sequences viewing behavior analysis using educational data-mining. Educ. Inf. Technol. (2021)

2. Kim, J., Guo, P.J., Cai, C.J., Li, S.W., Gajos, K.Z., Miller, R.C.: Data-driven interaction techniques for improving navigation of educational videos. In: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, pp. 563–572 (2014)
3. Sharma, A.K., Gupta, P.C., Kota, R.I.: Identifying the number of visitors to improve website usability from educational institution web log data. Int. J. Comput. Appl. Technol. Res. **2**(1), 22–26 (2012)
4. Krieter, P.: Can I record your screen? Mobile screen recordings as a long-term data source for user studies. In: Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia, pp. 1–10 (2019)
5. Alginahi, Y.: Preprocessing techniques in character recognition. Character Recognit. **1**, 1–19 (2010)
6. Heidarysafa, M., Reed, J., Kowsari, K., Leviton, A.C.R., Warren, J.I., Brown, D.E.: From videos to urls: a multi-browser guide to extract user's behavior with optical character recognition. In: Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), vol. 1, no. 1, pp. 503–514. Springer International Publishing (2020)
7. Menchen-Trevino, E.: Web historian: enabling multi-method and independent research with real-world web browsing history data. In: IConference 2016 Proceedings (2016)
8. Krieter, P., Breiter, A.: Analyzing mobile application usage: generating log files from mobile screen recordings. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 1–10 (2018)
9. Jadhav, S., Iyer, S., Arya, K.: Enhancing problem-solving in robotics education: analysis of scaffolds and interactions. In: 2024 IEEE International Conference on Advanced Learning Technologies (ICALT), pp. 121–123. Nicosia, North Cyprus, Cyprus, 2024 (2024)
10. OECD: PISA 2015 collaborative problem-solving framework, in PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving, OECD Publishing, Paris (2017)
11. Po, D.K.: Similarity based information retrieval using Levenshtein distance algorithm. Int. J. Adv. Sci. Res. Eng. **6**(04), 06–10 (2020)