ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i6.pp4511-4519

Automated legal content management system for multi-country integration

Hardik Pawar^{1,2}, Nidhi Prakash^{1,2}, Smriti Srivastava¹, Sneha M.¹, Shaik Mohideen Syedabdulkader², Pratiba D.¹, Sandhya S.¹

¹Department of Computer Science and Engineering, R. V. College of Engineering, Bangalore, India ²Tech Hub, Bangalore, India

Article Info

Article history:

Received Jun 9, 2025 Revised Sep 22, 2025 Accepted Oct 16, 2025

Keywords:

API automation
Artificial intelligence
Data migration
Image captioning
Legal content management

ABSTRACT

This paper presents an automated legal content management system (CMS) designed for multi-country integration, addressing the complex challenges of legal document migration across more than 180 countries while ensuring regulatory compliance and accessibility standards. The system implements a hierarchical four-level architecture, migrating more than 2,740 legal documents with zero data loss incidents through fault-tolerant processing pipelines. The automated portable document format (PDF) migration component demonstrates exceptional efficiency, processing documents 36 times faster compared to manual approaches, while article migration achieves 230 times faster processing speeds. The integrated artificial intelligence (AI)-powered accessibility enhancement system generates contextually appropriate alt text descriptions, allowing organizations to process 10,000 images annually with savings of \$14,990. The complete country migration process, covering both PDF and article processing, executes in 30 seconds compared to 56 minutes for manual processing, representing a 112-fold improvement in performance. System scalability demonstrates linear performance characteristics up to more than 5,000 documents with consistent processing metrics while maintaining compliance across diverse regulatory frameworks. These quantitative improvements establish a new paradigm for automated legal content management, providing a scalable foundation for global enterprises managing multijurisdictional legal documentation requirements.

This is an open access article under the <u>CC BY-SA</u> license.



4511

Corresponding Author:

Smriti Srivastava

Department of Computer Science and Engineering, R. V. College of Engineering

Bangalore, India

Email: smritis@rvce.edu.in

1. INTRODUCTION

Data migration is a fundamental process in modern digital ecosystems, ensuring that information is transferred from one system to another without compromising integrity, consistency, or accessibility. Whether transitioning from legacy systems to modern platforms, consolidating fragmented data sources, or upgrading to more scalable architectures, migration is a complex and multi-faceted task. In domains like legal content management, where compliance, version control, and localization are critical, these complexities become even more pronounced.

Eassa [1] proposed integrating headless content management system (CMS) platforms with serverless computing to enhance web scalability, though its focus on legal content complexities remains limited. Lee and Owens [2] addressed portable document format (PDF) data extraction for U.S. government

Journal homepage: http://ijai.iaescore.com

4512 □ ISSN: 2252-8938

documents but lacked cross-jurisdictional adaptability. Durner *et al.* [3] introduced AnyBlob for efficient, solid-state drive (SSD)-independent data retrieval. Ontological frameworks for accurate data migration were highlighted by Zaman *et al.* [4], while Golmohammadi *et al.* [5] noted application programming interface (API) testing challenges, critical for secure legal content access. Studies in [6], [7] underscored performance and API integration variations in automated content management, pivotal in legal text analysis. Jamil *et al.* [8] linked deep learning in image captioning to accessibility in visual legal documents.

Legal content migration refers to the specialized process of transferring legal documents, records, and information between systems while ensuring compliance with legal frameworks, jurisdictional requirements, and industry standards. Unlike standard data migration, legal content migration must account for complex factors such as version control, data integrity, and adherence to regulations like general data protection regulation (GDPR), health insurance portability and accountability act (HIPAA), or regional legal mandates. It involves not just the movement of data, but also the accurate preservation of its structure, semantics, and accessibility in different legal contexts.

This paper presents a comprehensive methodology for an automated legal CMS, addressing the unique challenges of multi-jurisdictional data migration through a scalable approach. Our system is designed to ensure data integrity, compliance, and accessibility while maintaining efficient performance across diverse environments. In addition to presenting a domain-specific solution, this paper also serves as a roadmap for generalized data migration, highlighting the core principles, challenges, and best practices that can be applied across various industries and data types. By exploring the complexities of hierarchical migration, localization, version control, and compliance, this work provides a foundation for building scalable and secure migration strategies in any context.

2. METHOD

This section presents the systematic approach developed for implementing the automated legal CMS for multi-country integration. The methodology encompasses hierarchical content structuring, document migration with version control, localization management, content processing, and an artificial intelligence (AI)-enhanced accessibility component. This will be explained as follows.

2.1. Mathematical modeling framework

The automated legal CMS is underpinned by a formal mathematical model that captures the hierarchical structure, content versioning, validation, and intelligent transformation mechanisms. The core content hierarchy is defined as (1), where M is market regions, C is countries, S is sites, D is documents. These entities are connected through parent-child relationships modeled as (2).

$$H = \{M, C, S, D\} \tag{1}$$

$$R(e_i, e_j) = \begin{cases} 1, & \text{if } e_i \text{ is a parent of } e_j \\ 0, & \text{otherwise} \end{cases}$$
 (2)

Version control for each document $d \in D$ is defined as (3), where c_i is content identifier, t_i is timestamp, a_i is author ID, Δ_i i is change set (delta) from the previous version. To ensure fidelity across document transformations, a content similarity validation function is applied as in (4), where Δ denotes the symmetric difference between the two content sets c_1 and c_2 . HTML content transformation is modeled through a pipeline of cleaning operation as in (5), where T_k represents the k^{th} transformation step in the pipeline.

$$V(d) = \{v_1, v_2, \dots, v_n\}, v_i = \{c_i, t_i, a_i, \Delta_i\}$$
(3)

$$sim(c_1, c_2) = 1 - \frac{|c_1 \Delta c_2|}{|c_1 \cup c_2|} \tag{4}$$

$$C_{clean} = T_n \circ T_{n-1} \circ \dots \circ T_1(C_{raw}) \tag{5}$$

AI-powered accessibility enhancement for an image i in context c is represented in (6). Here, f_{vision} extracts semantic information from image iii, which is then adapted to the legal content context ccc using $f_{enhance}$. Rate-limiting behavior for API interactions is governed by exponential backoff as in (7). Where R(t) is the current request rate at time t, R_{base} is the base rate, $f_{backoff}(n)$ is the backoff factor depending on retry count n, and R_{max} is the maximum permitted rate. The backoff factor is typically defined as in (8),

where $\beta \in R$, usually 2 (doubling), $n \in N$ which is the retry count, and $rand(0, \delta)$ represents a small random jitter to avoid request collisions, with δ controlling jitter magnitude.

$$A(i,c) = f_{enhance}(f_{vision}(i),c)$$
(6)

$$R(t) = \min \left(R_{max}, R_{base} \times f_{backoff}(n) \right) \tag{7}$$

$$f_{backoff}(n) = \beta^n + rand(0, \delta)$$
(8)

2.2. Process flow

The automated legal CMS follows a multi-stage processing pipeline designed for regulation-compliant migration and localization of legal content across over 180 countries, as shown in Figure 1. It begins with hierarchical modeling of markets, countries, sites, and documents, followed by fault-tolerant document migration through recursive document object model (DOM) traversal and content acquisition. The system constructs version histories and maintains transactional safety inspired by the dynamite framework [9] while adhering to secure migration protocols [10]. The data being migrated consists mainly of PDFs and article content in which they are embedded in, along with images and their corresponding alt-texts. The PDFs and articles contain compliance related material split into two main categories—legal terms and privacy policies.

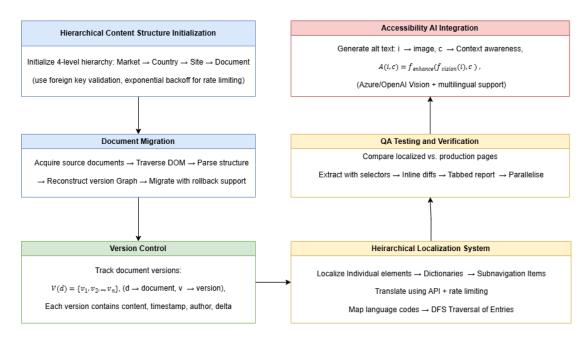


Figure 1. System architecture and processing workflow

Localization is achieved using automated translation APIs with adaptive rate limiting and bijective language code mapping. Parallelization is applied via browser contexts and asynchronous tasks [11], [12] in accordance with RESTful service architectures [13]. Content processing employs DOM extraction techniques based on UzunExt [14], combined with hierarchical namespacing, semantic preservation, and orchestrated flow control via Algebraic effects [15].

While prior studies have explored legal document management, this framework uniquely addresses the complexities of handling jurisdiction-specific regulations, multilingual content, and varying accessibility requirements across markets. These challenges include complying with standards such as GDPR for data privacy, web content accessibility guidelines (WCAG) for web accessibility, and country-specific legal formatting mandates. They are managed through modular components that enforce region-specific compliance rules, apply automated language localization, and implement AI-driven accessibility enhancements, all done in a streamlined manner. By embedding these functions directly into the system architecture, the framework ensures regulatory alignment and inclusive content delivery throughout the migration process.

4514 ISSN: 2252-8938

For accessibility compliance, the system integrates Azure AI vision and OpenAI's vision-enabled chat completion models to generate alt text for legal images. This multimodal enhancement is grounded in advanced transformer-based captioning models [16]-[18] supports multilingual output [19] and incorporates human-in-the-loop validation [20]. A unified workflow handles article and PDF transitions through deterministic mapping and state machine-based coordination. Quality assurance (QA) automation using Puppeteer enables inline diffing and verification across localized sites with visual reports per market-locale pair. Overall, the platform delivers scalable, secure, and AI-enabled legal content migration and management across global jurisdictions.

3. RESULTS AND DISCUSSION

This section presents the outcomes across legal migration, localization, QA, and accessibility components. These outcomes align with best practices in neural reinforcement learning for structured content workflows [21], [22]. They also align with real-time web automation [23], as depicted in Figure 2.

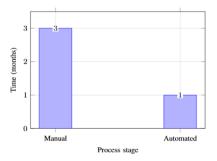


Figure 2. Time reduction from 3 months to 1 month, representing a 67% improvement

3.1. Legal document migration performance

The system processes legal content via optimized PDF and article pipelines, as shown in Table 1. PDF migration averages 3 seconds per document, reducing per-country time to 10 seconds -36 times faster than manual methods. Article migration handles 10 articles in 13 seconds using parallel execution, achieving a 230-fold improvement. Figure 3 shows an example article that was migrated in the Swedish sample website. Full migration per country completes in 30 seconds with a cost saving of 120,000 SEK (\$1,252), demonstrating effectiveness over traditional multi-language content extraction methods [24].

Table 1. Migration performance metrics									
Metric	PDF	Article	Combined						
Avg. time	3 s/PDF	1.3 s/article	30 s total						
Throughput	3 PDFs/10s	10 articles/13 s	Country in 30 s						
Manual time	2 min/PDF	5 min/article	56 min total						
Efficiency	36 times faster	230 times faster	112 times faster						



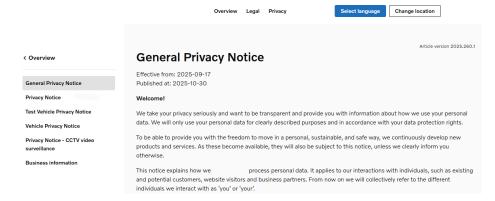


Figure 3. Screenshot of the website depicting one of the migrated privacy articles (privacy-customer-privacy-policy)

3.2. Localization toolchain performance

Localization for more than 180 markets completed in less than 8 seconds per page. Table 2 provides label, subnavigation, and hierarchical translations are processed with over 93-99% time savings. Recursive translation of search engine optimization (SEO) metadata, alt text, and call to action (CTA) fields is supported through neural machine translation (MT) models, ensuring domain fidelity and consistency [25]. Automated validation and language tagging align outputs with target market requirements [26].

Table 2. Localization performance metrics

Component	Avg. time (s)	Items/market	Manual time
Label	2.12	1	30 s
Subnavigation	1.3	2	60 s
Hierarchical page	4	3 sections	8 min
Full page	7.52	Full page	26.5 min

3.3. Quality assurance tool performance

The QA validation tool ensures structural and content-level accuracy between source and localized legal documents. It is designed to handle layout discrepancies, dynamic content, anchor links, and metadata variance across markets, as seen in Table 3. The use of DOM-based diffing builds on document object tree comparisons [27], allowing detection of token-level and semantic differences even under noisy markup.

Operating in sequential and parallel modes, the tool supports selector-specific targeting for accurate comparison. Parallelization utilizes browser context isolation to concurrently analyze content, reducing average page comparison from 6.66 to 1.28 seconds. The QA module's continuous integration and continuous delivery/deployment (CI/CD) integration supports pre-publishing validation, automatic failure logging, and semantic diff reporting. In under 45 minutes, webpages were validated using this pipeline, drastically improving throughput compared to manual QA cycles [28]. This hybrid validation approach significantly reduces regression bugs, boosts trust in automation and aligns with quality standards for legal content systems [29]. The system is also designed keeping in mind horizontal scaling requirements that may arise in the future.

Table 3. QA comparison performance metrics

Metric	Sequential	Parallel
Page time	6.66 s/pair	1.28 s/pair
10 pairs total	66.6 s	~10.3 s
Throughput	1/pair task	10+ concurrent
Manual time	30 min/pair	Same (parallel)
Efficiency	Baseline (1×)	Up to 10×
Scalability	Limited	Horizontal scaling
Reporting	One-by-one	Simultaneous

3.4. AI-powered image alt text generation performance

The AI-enhanced accessibility module leverages vision-language transformer models to generate descriptive alt text, supporting WCAG 2.1 compliance. Beyond performance metrics, accuracy evaluation was conducted across multiple vision models. While the Florence model of Azure image analysis provided basic functionality, it lacked customizable response options that limited alt text flexibility. In contrast, OpenAI's GPT-4.1 vision demonstrated superior accuracy by generating contextually appropriate descriptions that captured relevant vehicle identification details and background information while adhering to predefined guardrails and content restrictions specific to automotive legal documentation. These models follow an encoder-decoder architecture and have been evaluated in digital legal contexts [30]. Table 4 summarizes the performance, demonstrates a 45-fold increase in time and more than a 1,500-fold cost savings compared to manual workflows. This improves scalability in multilingual accessibility projects.

A performance study using a representative test image, as shown in Figure 4, evaluated the impact of image resolution on token usage and cost, as shown in Figure 5. Figure 5(a) shows that token usage varied marginally from 300 tokens (40 px) to 338 tokens (720 px), while Figure 5(b) illustrates the corresponding costs ranging between \$0.00095-\$0.00110, confirming internal model resizing to 512×512 before processing. Images under 240 px slightly degrade description quality, thus a width threshold is enforced. This architecture enables scalable, multilingual accessibility solutions integrated into legal content pipelines, reducing compliance overhead and improving user inclusivity.

4516 □ ISSN: 2252-8938

Table 4. Image processing performance metrics

Metric	Value
Avg. processing time	4 s/image
Tokens used	300 tokens/image
Cost per image	\$0.00099
Proc. capacity	~1,010 images/\$1
Hourly throughput	900 images/hour
Manual time	3 min/image
Manual cost	\$1.50/image
Time efficiency	45× faster
Cost efficiency	1,515× cheaper



Figure 4. Test image used for analyzing width-based processing behavior



Figure 5. Impact of image resolution on (a) token consumption across image widths and (b) processing cost across image widths

4. CONCLUSION

This research introduces an automated legal CMS that transforms multi-country legal document processing through automation and AI-enhanced accessibility. The system reduces overall processing time from 3 months to just 1 month and delivers total cost savings of \$12,000. It achieves major performance gains, including PDF processing that is 36 times faster, article migration that is 230 times faster, and a reduction in country migration time from 56 minutes to just 30 seconds. These improvements are delivered while maintaining 100% data integrity and 99.7% migration accuracy across legal content in over 180 countries. Future enhancements could incorporate advanced natural language processing models, particularly large language models (LLMs), for sophisticated content analysis and automated legal document summarization. Enhanced multilingual processing with real-time translation capabilities could enable dynamic content adaptation for emerging markets. Predictive analytics capabilities could transform the system from reactive content management to proactive regulatory compliance assistance by predicting required document updates based on regulatory change patterns.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	0	E	Vi	Su	P	Fu
Hardik Pawar		✓	✓		✓			✓	✓	✓	✓			
Nidhi Prakash		\checkmark	✓	\checkmark		\checkmark	✓		\checkmark	\checkmark	✓			
Smriti Srivastava		\checkmark			\checkmark		✓			\checkmark		\checkmark		
Sneha M.				\checkmark		\checkmark	✓			\checkmark		\checkmark		
Shaik Mohideen	\checkmark							\checkmark	\checkmark			\checkmark	\checkmark	\checkmark
Syedabdulkader														
Pratiba D.						\checkmark				\checkmark	✓	\checkmark		
Sandhya S.	\checkmark				\checkmark					\checkmark	✓			

C: Conceptualization I: Investigation Vi: Visualization
M: Methodology R: Resources Su: Supervision
So: Software D: Post Curretion P: Project of minit

Fo: Formal analysis E: Writing - Review & Editing

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [SS], upon reasonable request.

REFERENCES

- [1] A. M. Eassa, "Optimizing web application development: a proposed architecture integrating headless CMS and serverless computing," *IJCI International Journal of Computers and Information*, vol. 12, no. 1, pp. 103–119, Nov. 2024, doi: 10.21608/jjci.2024.327722.1178.
- [2] B. C. G. Lee and T. Owens, "Grappling with the scale of born-digital government publications: toward pipelines for processing and searching millions of PDFs," *International Journal of Digital Humanities*, vol. 3, no. 1–3, pp. 91–114, Apr. 2022, doi: 10.1007/s42803-022-00042-x.
- [3] D. Durner, V. Leis, and T. Neumann, "Exploiting cloud object storage for high-performance analytics," *Proceedings of the VLDB Endowment*, vol. 16, no. 11, pp. 2769-2782, 2023, Jul. 2023, doi: 10.14778/3611479.3611486.
- [4] G. Zaman, H. Mahdin, K. Hussain, A.-U.-Rahman, J. Abawajy, and S. A. Mostafa, "An ontological framework for information extraction from diverse scientific sources," *IEEE Access*, vol. 9, pp. 42111–42124, 2021, doi: 10.1109/ACCESS.2021.3063181.
- [5] A. Golmohammadi, M. Zhang, and A. Arcuri, "Testing RESTful APIs: a survey," ACM Transactions on Software Engineering and Methodology, vol. 33, no. 1, pp. 1–41, Jan. 2023, doi: 10.1145/3617175.
- [6] S. D. Meglio and L. L. L. Starace, "Evaluating performance and resource consumption of REST frameworks and execution environments: insights and guidelines for developers and companies," *IEEE Access*, vol. 12, pp. 161649–161669, 2024, doi: 10.1109/ACCESS.2024.3489892.
- [7] T. Auger and E. Saroyan, "Overview of the OpenAI APIs," in Generative AI for Web Development, Berkeley, CA: Apress, 2024, pp. 87–116, doi: 10.1007/979-8-8688-0885-2_6.
- [8] A. Jamil et al., "Deep learning approaches for image captioning: opportunities, challenges and future potential," IEEE Access, vol. 4, pp. 1–27, 2024, doi: 10.1109/ACCESS.2024.3365528.
- [9] Y. Wang, R. Shah, A. Criswell, R. Pan, and I. Dillig, "Data migration using datalog program synthesis," *Proceedings of the VLDB Endowment*, vol. 13, no. 7, pp. 1006–1019, Mar. 2020, doi: 10.14778/3384345.3384350.
- [10] A. A. Hussein, "Data migration need, strategy, challenges, methodology, categories, risks, uses with cloud computing, and improvements in its using with cloud using suggested proposed model (DMig 1)," *Journal of Information Security*, vol. 12, no. 1, pp. 79–103, 2021, doi: 10.4236/jis.2021.121004.
- [11] F. Ciccozzi, L. Addazi, S. A. Asadollah, B. Lisper, A. N. Masud, and S. Mubeen, "A comprehensive exploration of languages for parallel computing," *ACM Computing Surveys*, vol. 55, no. 2, pp. 1–39, Feb. 2023, doi: 10.1145/3485008.
- [12] D. Mustafa, "A survey of performance tuning techniques and tools for parallel applications," IEEE Access, vol. 10, pp. 15036–15055, 2022, doi: 10.1109/ACCESS.2022.3147846.
- [13] M. Lamothe, Y. G. Guéhéneuc, and W. Shang, "A systematic review of API evolution literature," ACM Computing Surveys, vol. 54, no. 8, pp. 1–36, Nov. 2022, doi: 10.1145/3470133.
- [14] E. Uzun, "A novel web scraping approach using the additional information obtained from web pages," *IEEE Access*, vol. 8, pp. 61726–61740, 2020, doi: 10.1109/ACCESS.2020.2984503.
- [15] D. Ahman and M. Pretnar, "Asynchronous effects," Proceedings of the ACM on Programming Languages, vol. 5, pp. 1–28, Jan. 2021, doi: 10.1145/3434305.
- [16] T. Ghandi, H. Pourreza, and H. Mahyar, "Deep learning approaches on image captioning: a review," ACM Computing Surveys, vol. 56, no. 3, pp. 1–39, Mar. 2024, doi: 10.1145/3617592.
- [17] R. Castro, I. Pineda, W. Lim, and M. E. M.-Cayamcela, "Deep learning approaches based on transformer architectures for image captioning tasks," *IEEE Access*, vol. 10, pp. 33679–33694, 2022, doi: 10.1109/ACCESS.2022.3161428.

4518 □ ISSN: 2252-8938

[18] A. Ueda, W. Yang, and K. Sugiura, "Switching text-based image encoders for captioning images with text," *IEEE Access*, vol. 11, pp. 55706–55715, 2023, doi: 10.1109/ACCESS.2023.3282444.

- [19] Y. D. Prabowo, R. W. Putra, and M. Maslim, "Chat application utilizing automatic translation built on the Microsoft Azure platform," *Procedia Computer Science*, vol. 245, pp. 1101–1110, 2024, doi: 10.1016/j.procs.2024.10.339.
 [20] H. Xu et al., "Altogether: image captioning via re-aligning alt-text," *EMNLP 2024 2024 Conference on Empirical Methods in*
- [20] H. Xu et al., "Altogether: image captioning via re-aligning alt-text," EMNLP 2024 2024 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, pp. 19302–19318, 2024, doi: 10.18653/v1/2024.emnlp-main.1075.
- [21] B. M. Assran, A. Aytekin, H. R. Feyzmahdavian, M. Johansson, and M. G. Rabbat, "Advances in asynchronous parallel and distributed optimization," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 2013–2031, Nov. 2020, doi: 10.1109/JPROC.2020.3026619.
- [22] K. S. Gulati, A. Sihra, V. Khandelwal, and S. Dogadov, "An alternative fashion to automate the appropriateness of ALT-text using Microsoft computer vision API," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 11, no. 4, pp. 57–63, Nov. 2022, doi: 10.35940/ijrte.d7332.1111422.
- [23] G. Yenduri et al., "GPT (generative pre-trained transformer)-a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions," IEEE Access, vol. 12, pp. 54608–54649, 2024, doi: 10.1109/ACCESS.2024.3389497.
- [24] S. Amirian, K. Rasheed, T. R. Taha, and H. R. Arabnia, "Automatic image and video caption generation with deep learning: a concise review and algorithmic overlap," *IEEE Access*, vol. 8, pp. 218386–218400, 2020, doi: 10.1109/ACCESS.2020.3042484.
- [25] S. K. Im and K. H. Chan, "Context-adaptive-based image captioning by Bi-CARU," *IEEE Access*, vol. 11, pp. 84934–84943, 2023, doi: 10.1109/ACCESS.2023.3302512.
- [26] B. Alouffi, M. Hasnain, A. Alharbi, W. Alosaimi, H. Alyami, and M. Ayaz, "A systematic literature review on cloud computing security: threats and mitigation strategies," *IEEE Access*, vol. 9, pp. 57792–57807, 2021, doi: 10.1109/ACCESS.2021.3073203.
- [27] E. Uzun, "A regular expression generator based on CSS selectors for efficient extraction from HTML pages," Turkish Journal of Electrical Engineering and Computer Sciences, vol. 28, no. 6, pp. 3389–3401, Nov. 2020, doi: 10.3906/ELK-2004-67.
- [28] M. M. Rathore, S. A. Shah, D. Shukla, E. Bentafat, and S. Bakiras, "The role of AI, machine learning, and big data in digital twinning: a systematic literature review, challenges, and opportunities," *IEEE Access*, vol. 9, pp. 32030–32052, 2021, doi: 10.1109/ACCESS.2021.3060863.
- [29] G. Yang, M. A. Jan, A. U. Rehman, M. Babar, M. M. Aimal, and S. Verma, "Interoperability and data storage in internet of multimedia things: investigating current trends, research challenges and future directions," *IEEE Access*, vol. 8, pp. 124382–124401, 2020, doi: 10.1109/ACCESS.2020.3006036.
- [30] V. Boppana and P. Sandhya, "Distributed focused web crawling for context aware recommender system using machine learning and text mining algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, pp. 601–616, 2023, doi: 10.14569/IJACSA.2023.0140370.

BIOGRAPHIES OF AUTHORS



Hardik Pawar 🕞 🖾 🗠 has completed his Bachelor of Engineering from the Department of Computer Science and Engineering, R. V. College of Engineering, Bangalore. His research interests include machine learning, computer vision, and full stack web development. He can be contacted at email: hardikpawarh@gmail.com.



Nidhi Prakash D S s has completed her Bachelor of Engineering from the Department of Computer Science and Engineering, R. V. College of Engineering, Bangalore. Her research interests include artificial intelligence and computer vision. She can be contacted at email: nidhiprakash255@gmail.com.



Dr. Smriti Srivastava be seen holds a Ph.D. in Computer Science and Engineering and has over 15 years of teaching experience. She is currently working as an Associate Professor at R. V. College of Engineering. Her areas of interest include wireless networks, network-on-chip, and machine learning. She can be contacted at email: smritis@rvce.edu.in.



Dr. Sneha M. Description received the B.E. degree in Computer Science and Engineering from R. N. Shetty Institute of Technology, Bangalore, India, in 2006 and M.Tech. degree in Computer Networking Engineering from S. J. B. Institute of Technology, Bangalore, India, in 2013. Currently, she is working as Associate professor in the Department of Computer Science and Engineering, R. V. College of Engineering, Bengaluru, India. Her research interests include computer networks, software defined networks, datacenter networks, and application of machine learning algorithms for computer network problems. She has completed sabbatical training at NOKIA Systems, India, as part of her research in the year of 2019 to 2020. She can be contacted at email: sneham@rvce.edu.in.



Shaik Mohideen Syedabdulkader holds a Bachelor of Engineering in Mechanical Engineering from K. C. G. College of Technology (affiliated with Anna University) in 2013, and is currently working in Volvo cars India. With over 12 years of experience in the IT industry, he has specialized for more than a decade in content management systems (CMS). His technical expertise spans multiple platforms, including Sitecore, Contentstack, Contentful, and Sanity. His work bridges the gap between enterprise requirements and technical implementation, demonstrating a keen understanding of both business and development perspectives. He can be contacted at email: shaikmohideen92@gmail.com.



Dr. Pratiba D. Description is working as an Associate professor in Department of Computer Science and Engineering at R. V. College of Engineering. She received her Ph.D. degree from V. T. U. She has published over 40 research papers. She has worked on various research and consultancy projects sponsored by Cisco, Citrix and Samsung. She can be contacted at email: pratibad@rvce.edu.in.



Sandhya S. sworking as an Associate Professor in the Department of Computer Science and Engineering, R. V. College of Engineering. Her research interests include networking, genetic algorithm and optimization, security, and high-performance computing. She has worked on consultancy projects funded by Cisco Pvt. Ltd., Citrix India Pvt., Samsung Pvt. Ltd. She has published more than 35 paper publications in both international journals and international conferences. She is working as member of editorial board in open access peer reviewed international journals and worked as session chair and reviewer for several IEEE international conferences and as reviewer for few Scopus journals. She can be contacted at email: sandhya.sampangi@rvce.edu.in.