IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

An Effective Semantic Code Clone Detection Frame Work using Pairwise Feature Fusion

Dr. Nilabar Nisha, Arish S, Dinesh kumar P, Hariharan V, Sakthivel E
Department of Computer Science and Engineering
Mahendra Institute of Engineering and Technology, Salem, India

Abstract: *This paper presents the design, implementation, and evaluation of an advanced website cloning tool developed to address the growing need for efficient web archiving solutions. The tool enables users to create local copies of websites with their original structure and assets intact, supporting various use cases including offline access, web development, digital preservation, and comparative analysis. Through a systematic approach to web crawling, content extraction, and resource management, the system offers configurable crawling depths, selective asset downloading, and support for dynamic content rendering. The implementation leverages modern web technologies including Next.js, React, and Node.js to create a responsive and intuitive user interface. Evaluation results demonstrate the tool's effectiveness in accurately cloning diverse websites while maintaining performance and scalability. This paper contributes to the field of web archiving by providing insights into the technical challenges and solutions for comprehensive website preservation in an increasingly complex web ecosystem.*

Keywords: Artificial Intelligence

I. INTRODUCTION

The World Wide Web has evolved into a vast repository of human knowledge, culture, and digital artifacts since its inception in the early 1990s. Websites, as the
primary interface for accessing this information, represent a significant portion of our collective digital heritage. However, the ephemeral nature of web content poses a substantial challenge for preservation efforts. Websites frequently change, become unavailable, or disappear entirely, resulting in what is commonly referred to as "link rot" and "content drift" (Zittrain et al., 2014).

II. LITERATURE REVIEW

Web archiving emerged as a field of practice and research in the mid-1990s, coinciding with the rapid growth of the World Wide Web. The Internet Archive, founded by Brewster Kahle in 1996, pioneered large-scale web archiving with its Wayback Machine, which has since become the largest publicly accessible web archive (Kahle, 1997). National libraries and archives soon followed, with institutions such as the National Library of Australia (PANDORA Archive), the British Library (UK Web Archive), and the Library of Congress establishing their own web archiving programs.

III. RESEARCH METHODOLOGY

Web archiving emerged as a field of practice and research in the mid-1990s, coinciding with the rapid growth of the World Wide Web. The Internet Archive, founded by Brewster Kahle in 1996, pioneered large-scale web archiving with its Wayback Machine, which has since become the largest publicly accessible web archive (Kahle, 1997). National libraries and archives soon followed, with institutions such as the National Library of Australia (PANDORA Archive), the British Library (UK Web Archive), and the Library of Congress establishing their own web archiving programs.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

IV. SYSTEM ARCHITECTURE

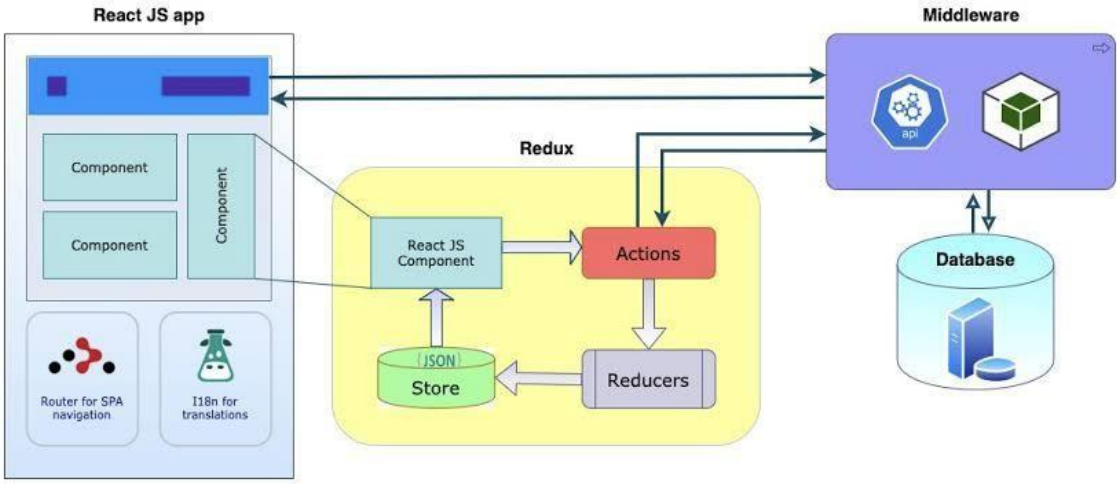

Fig 1. System architecture model

V. EXPERIMENTAL RESULTS

Web archiving emerged as a field of practice and research in the mid-1990s, coinciding with the rapid growth of the World Wide Web. The Internet Archive, founded by Brewster Kahle in 1996, pioneered large-scale web archiving with its Wayback Machine, which has since become the largest publicly accessible web archive (Kahle, 1997). National libraries and archives soon followed, with institutions such as the National Library of Australia (PANDORA Archive), the British Library (UK Web Archive), and the Library of Congress establishing their own web archiving programs.

VI. CONCULSION AND FUTURE WORK

This research has made several significant contributions to the field of web archiving and digital preservation through the design, implementation, and evaluation of a comprehensive website. cloning tool The research has demonstrated the effectiveness of an. integrated approach to website archiving that combines traditional crawling techniques with dynamic content rendering. This integration addresses one of the most significant challenges inmodern.web archiving: the preservation of JavaScript-dependent content and single-page, applications.

REFERENCES

[1]. Besek, J. M. (2003). Copyright issues relevant to the creation of a digital archive: A preliminary assessment. Council on Library and Information Resources.
[2]. Brunelle, J. F., Kelly, M., Weigle, M. C., & Nelson, M. L. (2014). The impact of JavaScript on archivability. International Journal on Digital Libraries, 15(2-4), 239- 252.
[3]. Brunelle, J. F., Kelly, M., SalahEldeen, H., Weigle, M. C., & Nelson, M. L. (2015). Not all mementos arecreated equal: Measuring the impact of missing resources.
International Journal on Digital Libraries,16(3-4), 283-301.
[4]. Brunelle, J. F., Kelly, M., Weigle, M. C., & Nelson, M. L. (2016). The impact of JavaScript onarchivability. International Journal on Digital Libraries, 17(2), 95-117.
[5]. Giles, C. L., Sun, Y., & Councill, I. G. (2010). Measuring the web crawler ethics. In Proceedings of the19th international conference on World wide web (pp. 1101- 1102).
[6]. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research.MIS quarterly, 75-105.
[7]. Kahle, B. (1997). Preserving the internet. Scientific American, 276(3), 82-83.
Masanès, J. (2006). Web archiving methods and approaches: A comparative study. Library Trends,54(1), 72-90.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

[8]. Millard, D. E., Hargood, C., Jewell, M. O., & Weal, M. J. (2013). Canyons, deltas and plains: Towards aunified sculptural model of location-based hypertext. In Proceedings of the 24th ACM Conference onHypertext and Social Media (pp. 109- 118).
[9]. Mohr, G., Kimpton, M., Stack, M., & Ranitovic, I. (2004). Introduction to Heritrix, an archival quality web crawler. In Proceedings of the 4th International Web Archiving Workshop. Nikšić, H. (2005). GNU Wget. Retrieved from https://www.gnu.org/software/wget/