Evaluasi Strategi Context-Aware Large Language Model (LLM) untuk Pembuatan Kode Pengujian Otomatis
pada Framework Playwright
Risma Saputri, Divi Galih Prasetyo Putri, S.Kom., M.Kom., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

EVALUASI STRATEGI *CONTEXT-AWARE LARGE LANGUAGE MODEL* (LLM) UNTUK PEMBUATAN KODE PENGUJIAN OTOMATIS PADA *FRAMEWORK* PLAYWRIGHT

Risma Saputri

22/505613/SV/21835

Pembuatan skrip pengujian otomatis menggunakan *Large Language Model* (LLM) menjadi solusi potensial untuk menjembatani *semantic gap* antara dokumentasi kebutuhan dan implementasi teknis skrip. Akan tetapi, LLM tanpa konteks tambahan cenderung mengalami halusinasi terhadap elemen antarmuka yang mengakibatkan rendahnya tingkat keberhasilan eksekusi. Penelitian ini mengevaluasi efektivitas pemberian konteks HTML dan integrasi *Model Context Protocol* (MCP) Playwright dalam menghasilkan skrip pengujian otomatis berbasis Playwright. Penelitian dilakukan dengan membandingkan tiga strategi integrasi konteks terhadap *baseline* LLM. Eksperimen dilakukan terhadap 12 *test case* pada aplikasi AssistLab menggunakan model Claude Haiku 4.5 dengan tiga iterasi tiap kondisi eksperimen. Hasil penelitian menunjukkan bahwa pemberian konteks HTML meningkatkan *Execution Success Rate* sebesar 63,89 pp dari 13,89% menjadi 77,78%, dan peningkatan *Semantic Relevance* sebesar 6,45 pp dari 90,99% menjadi 97,43% dibandingkan *baseline* LLM. Peningkatan performa ini dicapai dengan *trade-off* penambahan *prompt size* sebesar 3,05K, tetapi diimbangi dengan penurunan latensi sebesar 282,22 detik. Penelitian ini menyimpulkan bahwa pemberian konteks HTML merupakan pendekatan paling efektif dan efisien dalam menghasilkan skrip pengujian otomatis menggunakan LLM.

Kata kunci : *Framework* Playwright, *Large Language Model, Model Context Protocol,* Pembuatan Skrip Pengujian, Pengujian Otomatis

UNIVERSITAS
GADJAH MADA
**Evaluasi Strategi Context-Aware Large Language Model (LLM) untuk Pembuatan Kode Pengujian Otomatis
pada Framework Playwright**
Risma Saputri, Divi Galih Prasetyo Putri, S.Kom., M.Kom., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

EVALUATION OF CONTEXT-AWARE LARGE LANGUAGE MODEL (LLM) STRATEGIES FOR AUTOMATED TEST CODE GENERATION USING THE PLAYWRIGHT FRAMEWORK

Risma Saputri

22/505613/SV/21835

Automated test script generation using Large Language Models (LLM) has emerged as a potential solution to bridge the semantic gap between requirements documentation and technical script implementation. However, LLMs without additional context tend to hallucinate interface elements, resulting in low execution success rates. This study evaluates the effectiveness of providing HTML context and integrating the Model Context Protocol (MCP) Playwright in generating automated test scripts based on Playwright. The study was conducted by comparing three context integration strategies against an LLM baseline. Experiments were performed on 12 test cases of the AssistLab application using the Claude Haiku 4.5 model with three iterations for each experimental condition. The results indicate that providing HTML context improves the Execution Success Rate by 63.89 pp from 13.89% to 77.78%, and the Semantic Relevance by 6.45 pp from 90.99% to 97.43% compared to the LLM baseline. These performance improvements were achieved with a trade-off of a prompt size increase of 3.05K tokens, offset by a latency reduction of 282.22 seconds. This study concludes that providing HTML context is the most effective and efficient approach for generating automated test scripts using LLMs.

Keywords: Automation Testing, Large Language Model, Model Context Protocol, Playwright Framework, Test Script Generation