UNIVERSITAS
GADJAH MADA
Metode Lokalisasi Bug Berbasis Struktur Repositori pada GitHub Open-Source
Khairul Ihsannudin, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR ISI

Halaman Judul **ii**

Halaman Pengesahan **iii**

Halaman Persembahan **iv**

Halaman Motto **v**

PRAKATA **vi**

PERNYATAAN BEBAS PLAGIASI **xv**

INTISARI **xvi**

ABSTRACT **xvii**

I PENDAHULUAN **1**

1.1 Latar Belakang 1
1.2 Rumusan Masalah 5
1.3 Tujuan Penelitian 5
1.4 Manfaat Penelitian 6
1.5 Batasan Penelitian 6

II TINJAUAN PUSTAKA **7**

2.1 *Benchmark* Isu Repositori *Open Source* Github 7
2.2 *Bug Localization* berbasis LLM 8
2.3 Pendekatan Berbasis Graf 9
2.4 Pendekatan Berbasis Struktur Repositori 10
2.5 Navigasi Berbasis Ringkasan Hierarkis 10
2.6 Pendekatan Berbasis Ringkasan Multi-Agen 12
2.7 Posisi Penelitian 12

III LANDASAN TEORI **17**

3.1 *Bug Localization* 17
3.2 *Information Retrieval* 18

UNIVERSITAS
GADJAH MADA
Metode Lokalisasi Bug Berbasis Struktur Repositori pada GitHub Open-Source
Khairul Ihsannudin, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

3.3 *Large Language Model* . . . 18
3.4 *Semantic Scoring* . . . 19
3.5 *Abstract Syntax Tree* . . . 20
3.6 Peringkasan Hierarkis (*Hierarchical Summarization*) . . . 22
3.7 SWE-bench . . . 23
3.7.1 SWE-bench Lite . . . 24
3.8 Metrik Evaluasi . . . 25
3.8.1 Metrik Akurasi Lokalisasi . . . 26
3.8.2 Metrik Efisiensi Penggunaan Token . . . 28

IV METODE PENELITIAN 30
4.1 Deskripsi Penelitian . . . 30
4.1.1 Definisi Masalah . . . 30
4.2 Pengumpulan Data . . . 31
4.3 Pra-pemrosesan Hierarki . . . 33
4.3.1 Ringkasan Level Fungsi . . . 33
4.3.2 Ringkasan Level File . . . 35
4.3.3 Ringkasan Level Folder . . . 35
4.3.4 Contoh Peringkasan Hierarkis pada Struktur Nested . . . 36
4.3.5 Strategi *Incremental Update* Ringkasan . . . 38
4.3.6 Pra-pemrosesan Data untuk *Ground Truth* . . . 39
4.4 Tahap Inferensi . . . 39
4.4.1 Pencarian Tingkat Folder . . . 40
4.4.2 Pencarian Tingkat File . . . 41
4.4.3 Pencarian Tingkat Fungsi . . . 41
4.4.4 Penilaian dan Pemeringkatan Semantik Berbasis LLM . . . 41
4.4.5 Algoritma *Bug Localization* Berbasis Struktur . . . 42
4.5 Evaluasi . . . 44
4.5.1 Pengaturan Evaluasi . . . 44
4.5.2 Pendekatan Pembanding . . . 45
4.5.3 Metrik Evaluasi . . . 46

V IMPLEMENTASI PENELITIAN 48
5.1 *Research Environment* . . . 48
5.2 Inisialisasi Pustaka dan Dataset . . . 48
5.3 Struktur Data (*Dataclasses*) . . . 49

UNIVERSITAS
GADJAH MADA
Metode Lokalisasi Bug Berbasis Struktur Repositori pada GitHub Open-Source
Khairul Ihsannudin, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

5.4 Pra-pemrosesan Hierarki 50
5.4.1 Pra-pemrosesan AST 51
5.4.2 Peringkasan dari Hasil Pra-pemrosesan 52
5.4.3 Alur Peringkasan Fungsi, Kelas, dan File 53
5.4.4 Alur Peringkasan Folder 54
5.5 Mekanisme *Incremental Update* 56
5.5.1 Deteksi Perubahan File 56
5.5.2 Checkout Commit (`GitHelper.checkout_commit`) . . 56
5.5.3 Orkestrasi Peringkasan Inkremental 57
5.6 Integrasi API Model Bahasa 58
5.7 Alur Pencarian Secara *Top-Down* (Inferensi) 60
5.8 Implementasi Modul Evaluasi dan Metrik 62

VI HASIL DAN PEMBAHASAN 63
6.1 Analisis Hasil Peringkasan Kode 63
6.1.1 Analisis Efisiensi Peringkasan Inkremental (*Incremental Summarization*) 63
6.1.2 Analisis Waktu Eksekusi dan Konsumsi Token Pra-pemrosesan 65
6.1.3 Analisis Tingkat Kompresi Token 67
6.2 Evaluasi Hasil Inferensi Metode Usulan dan Agentless pada Skenario Model *Open-Source* 68
6.2.1 Evaluasi Tingkat File 68
6.2.2 Evaluasi Tingkat Fungsi 72
6.2.3 Analisis Keberhasilan Lokalisasi per Repositori 74
6.2.4 Analisis Kesenjangan Lokalisasi File-ke-Fungsi 77
6.2.5 Efisiensi Fase Inferensi 78
6.3 Perbandingan Hasil Lokalisasi dengan Publikasi Sebelumnya 79
6.3.1 Perbandingan Biaya Inferensi Lokalisasi (GPT-4o-mini vs. GPT-4o) 80

VIIKESIMPULAN DAN SARAN 82
7.1 Kesimpulan 82
7.2 Saran 83

UNIVERSITAS
GADJAH MADA
Metode Lokalisasi Bug Berbasis Struktur Repositori pada GitHub Open-Source
Khairul Ihsannudin, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR TABEL

2.1 Perbandingan berbagai pendekatan dalam kasus *bug localization* . . . 14

3.1 Distribusi file, folder, dan jumlah isu pada SWE-bench Lite 25

4.1 SWE-bench Lite: sqlfluff__sqlfluff-1625 31
4.2 Perbandingan Elemen Pendekatan Lokalisasi Bug 45

5.1 Spesifikasi Komputer 48

6.1 Data Empiris Efisiensi Peringkasan Inkremental pada Dataset SWE-bench Lite 64
6.2 Waktu Komputasi dan Penggunaan Token Pra-pemrosesan pada Dataset SWE-bench Lite 65
6.3 Rata-rata Konsumsi Token pada Peringkasan Basis dan Inkremental . 66
6.4 Perbandingan Ukuran Token *Source Code* dan Representasi Ringkasan 67
6.5 Hasil Evaluasi File Hit@k dan Metrik Pemeringkatan pada Tingkat File 69
6.6 Hasil Evaluasi Recall@k pada Tingkat File 70
6.7 Hasil Evaluasi Func Hit@k dan Metrik Pemeringkatan pada Tingkat Fungsi 73
6.8 Hasil Evaluasi Recall@k pada Tingkat Fungsi 74
6.9 Perbandingan Hit Rate File Hit@5 dan Fungsi Hit@5 per Repositori (Skenario Nemotron-3 Nano 30B) 76
6.10 Perbandingan Efisiensi Komputasi 78
6.11 Perbandingan Akurasi Lokalisasi File Hit@1 dan Func Hit@1 pada SWE-bench Lite 79
6.12 Perbandingan Estimasi Biaya Inferensi Lokalisasi Bug 80

UNIVERSITAS
GADJAH MADA
Metode Lokalisasi Bug Berbasis Struktur Repositori pada GitHub Open-Source
Khairul Ihsannudin, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR GAMBAR

4.1 Proses pra-pemrosesan berbasis *bottom-up summarization* 34
4.2 Struktur hierarkis repositori Django (contoh) 36
4.3 Ilustrasi pembaruan ringkasan secara bertingkat. 38
4.4 Proses inferensi *top-down* untuk *bug localization*. 40

6.1 Diagram Venn tumpang tindih keberhasilan lokalisasi file (File Hit@5) antara Metode Usulan dan *Agentless* pada skenario Llama 3.1 8B (lokal) dan Nemotron-3 Nano 30B (cloud). 71
6.2 Distribusi keberhasilan lokalisasi file (File Hit@5) per repositori antara Metode Usulan (Nvidia) dan *Agentless* (Nvidia). 75
6.3 Rasio keberhasilan lokalisasi fungsi pada file teridentifikasi. 77

UNIVERSITAS
GADJAH MADA
Metode Lokalisasi Bug Berbasis Struktur Repositori pada GitHub Open-Source
Khairul Ihsannudin, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR KODE

3.1 Contoh fungsi Python sederhana 21
3.2 Serialisasi struktur AST dari fungsi `add` 21
5.1 Inisialisasi pustaka dan pemuatan dataset 48
5.2 Definisi struktur data hierarkis fungsi, kelas, file, dan folder 49
5.3 Fungsi pra-pemrosesan AST repositori menggunakan LocalPreprocessor 51
5.4 Fungsi peringkasan dari JSON pra-pemrosesan 52
5.5 Alur orkestrasi pemanggilan LLM pada objek FileInfo pra-pemrosesan 53
5.6 Orkestrasi peringkasan folder secara bottom-up 54
5.7 Fungsi deteksi file yang berubah di antara dua commit 56
5.8 Fungsi checkout commit pada repositori 56
5.9 Logika orkestrasi pembaruan ringkasan inkremental 57
5.10 Fungsi pemanggilan API untuk Ollama, OpenRouter, dan OpenAI . . 58
5.11 Inisialisasi dan pencatatan statistik penggunaan token 59
5.12 Orkestrasi alur inferensi top-down dan traversal rekursif 60
5.13 Fungsi penghitungan metrik Hit@k, Recall@k, MRR, dan MAP . . . 62