Universidade Federal de Sergipe
Engenharia de Software II

Diagnóstico e Auditoria do Projeto LangExtract

Atividade 3 – Etapa 1

Membros:
José Fernando Bispo dos Santos – 202200014210
Valter Fabricio dos Santos – 202000066991
Raphael Ferreira Portella Bacelar - 202100045822

Professor: Dr. Glauco de Figueiredo Carneiro

Janeiro de 2026

SUMÁRIO

1. INTRODUÇÃO

A evolução contínua de sistemas de software exige mecanismos que garantam a qualidade e a estabilidade do código ao longo do tempo. Nesse contexto, a Integração Contínua (Continuous Integration – CI) surge como uma prática fundamental, permitindo que alterações no código sejam verificadas automaticamente, reduzindo riscos de regressão e facilitando a colaboração entre desenvolvedores.

Esta atividade tem como objetivo analisar a maturidade das práticas de CI de um projeto de software open source real, no qual a equipe atua como engenheiros de DevOps. A partir dessa análise, busca-se compreender como a automação influencia a manutenibilidade do sistema e a entrada de novos contribuidores no projeto.

---

2. DESCRIÇÃO DO PROJETO AVALIADO

O projeto analisado nesta atividade é o **LangExtract**, um projeto open source mantido pela Google e disponibilizado na plataforma GitHub.

Repositório: https://github.com/google/langextract

LangExtract é uma biblioteca Python que utiliza LLMs para extrair informações estruturadas de documentos de texto não estruturados com base em instruções definidas pelo usuário. Ela processa materiais como anotações clínicas ou relatórios, identificando e organizando detalhes-chave, ao mesmo tempo em que garante que os dados extraídos correspondam ao texto de origem.

---

3. METODOLOGIA DE INVESTIGAÇÃO

A investigação do projeto foi realizada por meio de uma análise exploratória do repositório no GitHub, com foco nas práticas de Integração Contínua adotadas. Foram observados os seguintes aspectos:

- Existência do diretório `.github/workflows/`;
- Análise dos arquivos YAML de workflows;
- Verificação da aba *Actions* do GitHub;

- Observação do histórico de Pull Requests

Essa abordagem permitiu mapear o fluxo atual de desenvolvimento e identificar o nível de automação presente no projeto.

---

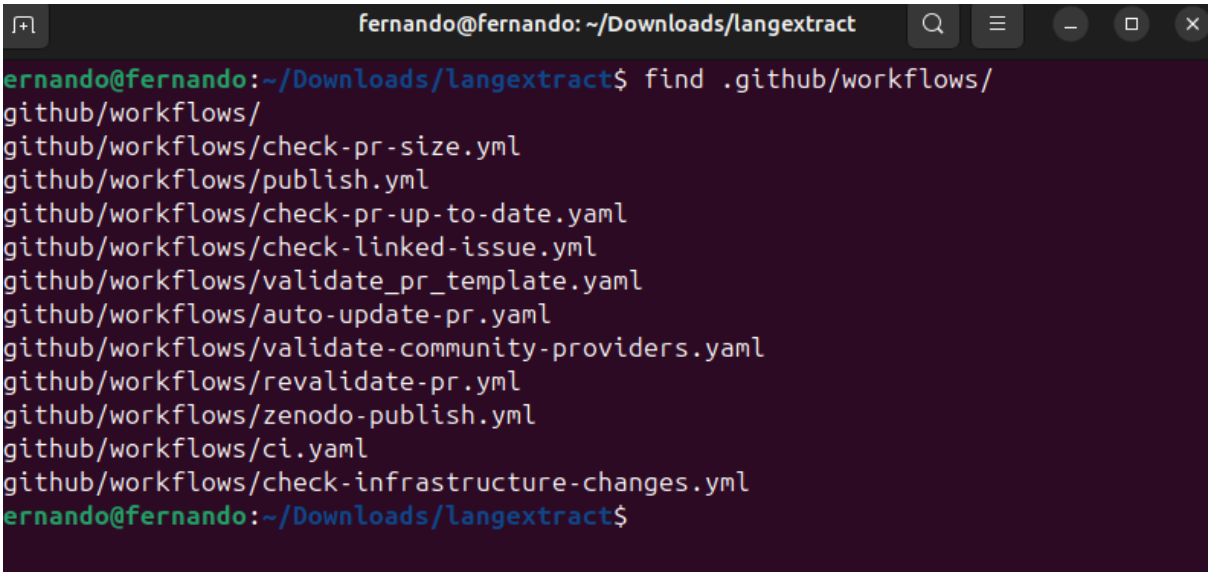
4. FERRAMENTAS DE CI/CD UTILIZADAS

O projeto **LangExtract utiliza o GitHub Actions como ferramenta de Integração Contínua (CI)**. Essa ferramenta é nativa da plataforma GitHub e permite a execução automática de tarefas em resposta a eventos como *push* e *pull request*.o

---

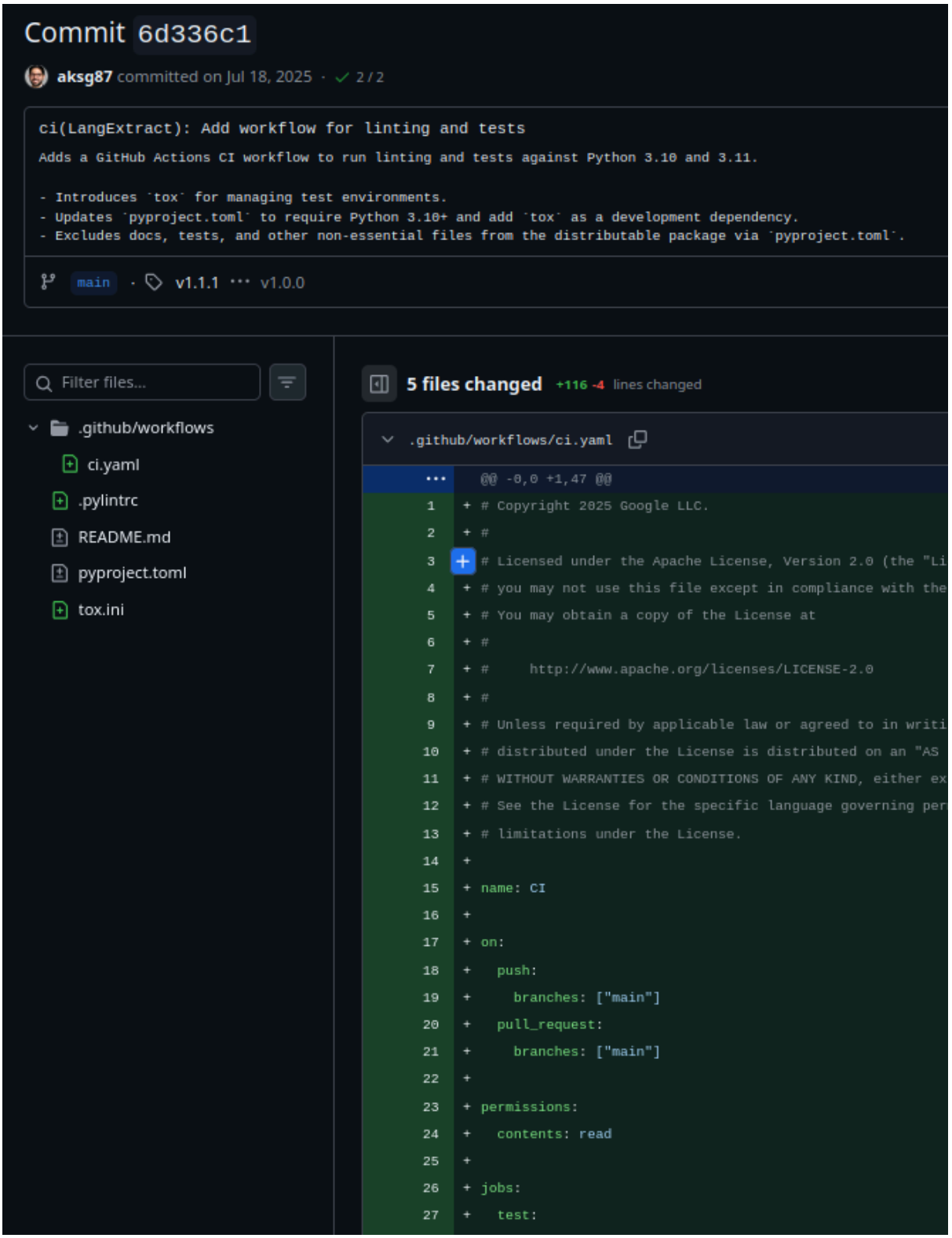
5. EVIDÊNCIAS TÉCNICAS

Como evidência da utilização de CI, foram identificados diversos arquivos de workflow no diretório `.github/workflows/`. Esses arquivos definem ações automatizadas que são executadas durante o ciclo de desenvolvimento do projeto.

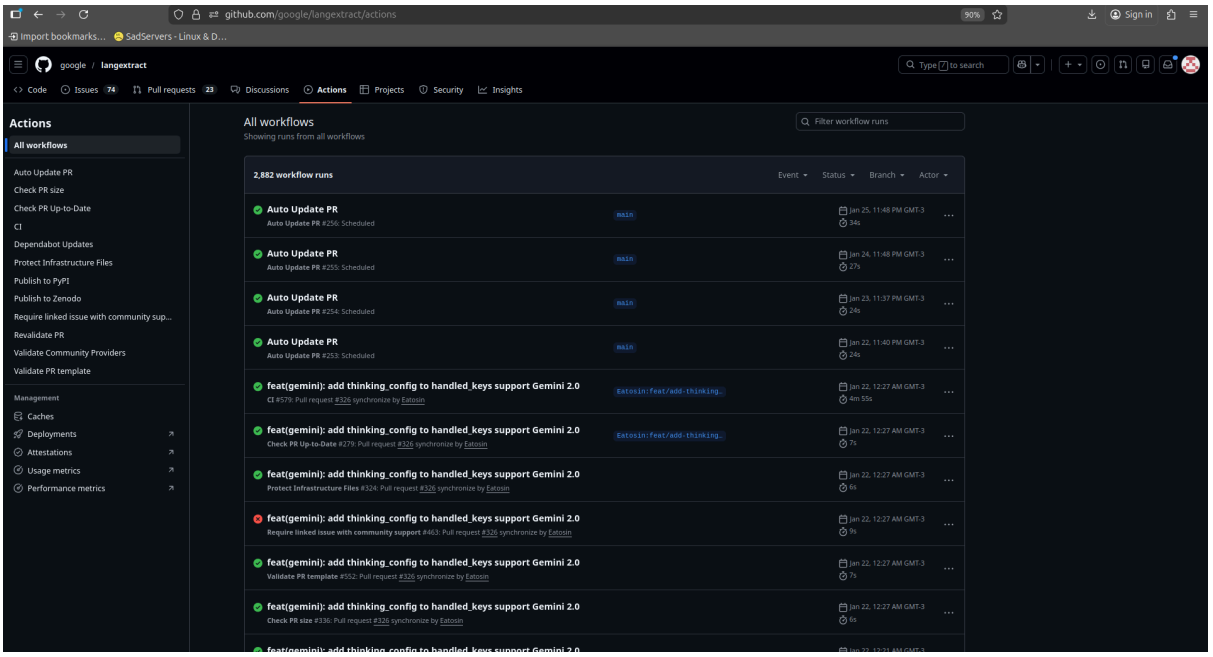


Complementarmente, foi realizada uma análise do histórico de commits, . Essa verificação permitiu identificar o primeiro commit diretamente relacionado à criação e evolução dos workflows de CI. O commit 6d336c1, no qual foi adicionada uma pipeline de GitHub Actions voltada para a execução de linting e testes automatizados, com suporte a múltiplas versões

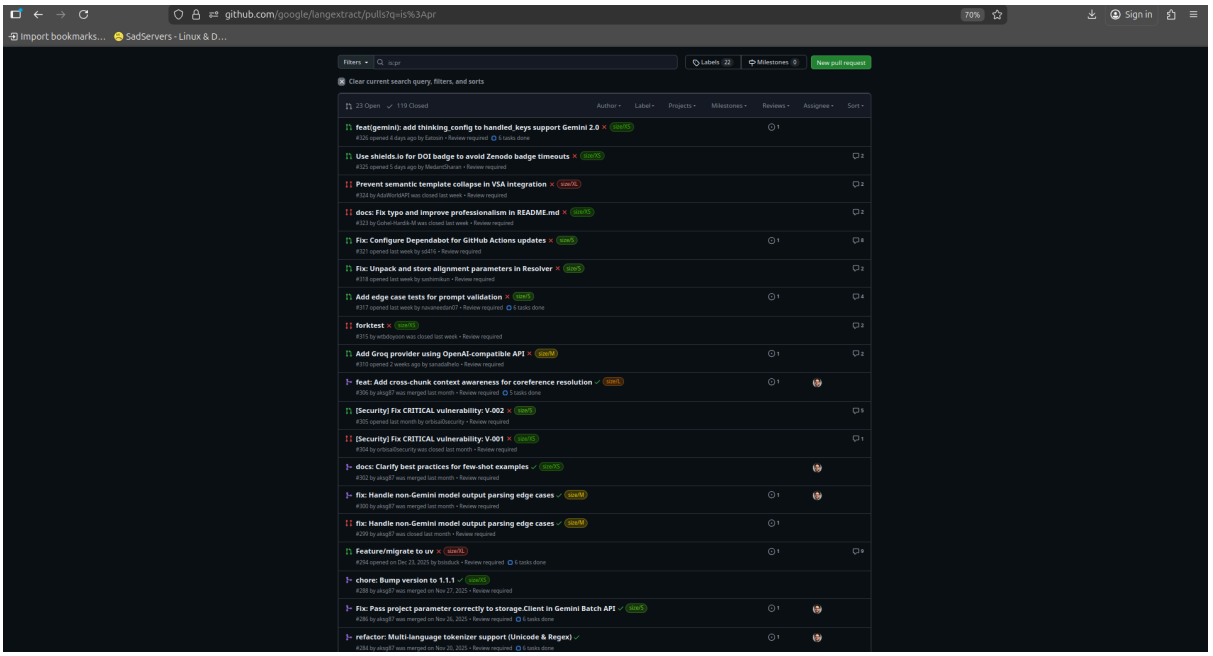
do Python (3.10 e 3.11), além da introdução da ferramenta tox para gerenciamento de ambientes de teste.



Além disso, a aba *Actions* do GitHub apresenta um histórico consistente de execuções dos workflows, indicando que os pipelines estão ativos e em uso contínuo.



No histórico de Pull Requests, é possível observar a presença de *status checks* ❌ ou ✅, que informam aos revisores se as verificações automatizadas foram concluídas com sucesso ou falha antes da integração do código.



Essas evidências confirmam que o projeto possui automação funcional e integrada ao processo de colaboração.

---

6. FLUXO ATUAL

Com base na análise realizada, o fluxo atual de desenvolvimento do projeto pode ser descrito da seguinte forma:

- Criação de alterações no código-fonte
    - Os contribuidores realizam modificações no código localmente e submetem essas alterações ao repositório por meio de commits.
- Abertura de Pull Requests (PRs)
    - As mudanças são integradas ao fluxo principal do projeto através da abertura de Pull Requests no GitHub, permitindo revisão por outros colaboradores ou mantenedores.
- Execução automática dos workflows de CI
    - Ao ocorrer um evento de *push* ou a abertura/atualização de um Pull Request, os workflows definidos no diretório `.github/workflows/` são acionados automaticamente pelo GitHub Actions.
- Feedback automatizado
    - Os resultados das execuções são exibidos diretamente no Pull Request por meio de *status checks* visuais (sucesso ou falha). Esse feedback imediato informa se o código atende aos critérios mínimos de qualidade antes da revisão humana.
- Revisão e integração do código
    - Apenas após a conclusão bem-sucedida das verificações automatizadas, o Pull Request pode ser revisado e, eventualmente, aprovado e mesclado ao branch principal do projeto.

---

7. CONCLUSÃO

A análise do projeto LangExtract demonstra um nível consistente de maturidade no uso de práticas de Integração Contínua. A adoção do GitHub Actions como ferramenta de CI, aliada à execução automatizada de linting e testes em múltiplas versões do Python, contribui diretamente para a manutenção da qualidade do código e para a detecção precoce de erros.

A presença de pipelines ativos e integrados ao fluxo de Pull Requests evidencia uma preocupação com a estabilidade do projeto e com a prevenção de regressões, fatores essenciais para a evolução sustentável do software. Além disso, o feedback automatizado fornecido pelas execuções de CI facilita a entrada de novos contribuidores, pois estabelece critérios claros e objetivos para a aceitação de mudanças.

Dessa forma, conclui-se que o LangExtract possui um processo de desenvolvimento bem estruturado no que se refere à Integração Contínua, reduzindo gargalos manuais e fortalecendo a confiabilidade do projeto ao longo de seu ciclo de vida.