

Relatório da Entrega Final Estudos Avançados em Sistemas de Software (05/07/2025)

Nome: David Tadokoro

NUSP: 10300507

Orientador: Paulo Meirelles

1) Contexto

No primeiro semestre de 2024, houve o oferecimento da disciplina de *Desenvolvimento em Software Livre* (MAC0470/5856), que envolveu mentorar os alunos no ecossistema do Linux, que inclui o próprio projeto do kernel Linux, ferramentas que suportam o desenvolvimento, distribuições GNU/Linux, entre outras. O programa da disciplina foi dividido em três fases, onde os alunos imergiram de forma prática em cada uma das camadas do ecossistema.

O projeto proposto para a disciplina envolve a escrita de um artigo científico sobre como treinar novos contribuidores de uma forma eficiente e prática com intuito de prepará-los com as habilidades necessárias para se tornarem reais desenvolvedores do kernel, baseado nas experiências da disciplina. Vale notar que o objetivo deste treinamento é de munir novos contribuidores (com pouca ou nenhuma experiência prévia em software livre) com a base essencial de habilidades e conhecimentos para irem além de serem os chamados *one-time contributors*.

O plano é que este artigo seja submetido ao ICSE 2026 - SEET.

2) Visão geral das entregas passadas

Proposta do Projeto A proposta feita ao início do semestre, em linhas gerais, permaneceu consistente e o cronograma foi respeitado ao decorrer da disciplina. Houve uma alteração pontual de não incluir os resultados do oferecimento da disciplina de Desenvolvimento em Software Livre de 2025, o que impactou os entregáveis a partir da entrega 4. No entanto, esta alteração manteve a mesma estrutura e objetivo do projeto (artigo) proposto inicialmente.

Entrega 1 O objetivo era organizar e analisar os dados advindos do questionário final da disciplina (no oferecimento de 2024). Alguns gráficos foram produzidos, mas o mais importante foi a análise qualitativa dos dados que (para esta fonte de dados) já começou a caracterizar os resultados do estudo.

Entrega 2 Seguindo a linha da primeira entrega, o objetivo desta era analisar os dados advindos das outras duas fontes de dados, que eram as observações dos mentores e os blogposts dos estudantes. A análise destas outras duas fontes foi mais qualitativa e, no caso dos blogposts, exigiu mineração e filtragem, mas, ao

final da entrega, a união das análises das duas entregas deu corpo à direção que iríamos tomar na confecção do artigo.

Entrega 3 A partir desta entrega (exceto a 7), todos os entregáveis eram seções do artigo. Para esta entrega 3, uma versão bem bruta e inicial do abstract e da seção de introdução foram produzidos. Neste momento, a introdução não possui embasamento teórico, algo que foi corrigido em certa medida nas seguintes entregas.

Entrega 4 Nesta entrega, o planejado era a seção de matérias e métodos, o que foi cumprido. Nesta entrega que houve a decisão em conjunto de reajustar do cronograma.

Entrega 5 Nesta entrega, o planejado era a seção de resultados, o que foi cumprido.

Entrega 6 Nesta entrega, o planejado eram as seções de discussão e conclusão, o que foi cumprido.

Entrega 7 Nesta penúltima entrega, o planejado era uma versão bruta juntando todas as partes do artigo com focando mais em coesão do que refino. Nesta etapa, tivemos uma primeira iteração do artigo, apesar de incompleta e bruta.

3) Entregável final planejado

Nesta entrega final, o objetivo era fazer uma primeira versão base para a submissão.

O arquivo fonte em Markdown (`entrega-final/relatorio.md`) deste relatório, se encontra no **meu repositório de artefatos para a disciplina**.

Como o artigo fechado depende de vários artefatos $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$, de forma que ficaria inviável colocar todo o texto neste relatório (como fizemos com as outras entregas para as seções), optamos por apenas referenciar o **documento Latex no overleaf para consulta**, que deve estar visível sem necessidade de liberação de acesso. Nele, além de ser possível consultar fonte do projeto, há o arquivo `paper.pdf` com o compilado do artigo.

4) Comentário sobre reajuste no cronograma

O reajuste no cronograma, o qual reduzia o escopo do artigo para apenas a edição de 2024 da disciplina (o plano original previa incluir a edição de 2025), foi motivado pelos seguintes fatores:

1. Originalmente, a inclusão dos dados de 2025 considerava que o artigo iria fechar o escopo apenas para o desenvolvimento kernel Linux, o que permitiria coletarmos os dados no meio do semestre a tempo de agregá-los

no artigo. Como determinamos que iríamos focar na experiência completa da disciplina, coletar e incluir estes dados tornou-se inviável.

2. Como um agravante, neste semestre, eu e o Paulo focamos (principalmente na segunda metade do semestre) na minha produção científica que era nula, o que resultou na divisão de esforços em relação à escrita dos artigos (incluindo o da disciplina). Tivemos 5 artigos submetidos e aceitos, sendo que, em 3 deles, consto como primeiro autor, fora mais 2 que foram submetidos ao VEM (em um deles sou primeiro autor). A seguir, a lista das publicações aceitas:

- D. Tadokoro, R. Siqueira, and P. Meirelles, "*Kworkflow: a Linux kernel Developer Automation Workflow System*" in *39th Brazilian Symposium on Software Engineering - Tools Track*.
- D. Tadokoro and P. Meirelles, "*Can the Linux kernel sustain 30 more years of growth? Toward mitigating bottlenecks in its development model*" in *39th Brazilian Symposium on Software Engineering - Insightful Ideas and Emerging Results Track*.
- D. Tadokoro, R. Passos, and P. Meirelles, "*Guidelines for Boosting Long-Lasting FLOSS Contributors*" in *DebConf 2025 - Academic Track*.
- R. Passos, A. Pilone, D. Tadokoro, and P. Meirelles, "*Streamlining Analysis on the Linux project with DUKS*" in *"VISSOFT 2025 - Visualization Challenge"*.
- L. Arcanjo, D. Tadokoro, and P. Meirelles, "*ArKanjo: a tool for detecting function-level Code Duplication in the Linux Kernel*" in *DebConf 2025 - Academic Track*.

5) Conclusão da Entrega Final

Com esta entrega final, produzimos uma segunda iteração do artigo que resolve a maioria das questões que foram levantadas ao decorrer das entregas (principalmente nas seções de conclusão dos relatórios). Apesar deste grande avanço, o artigo ainda não se encontra pronto para a submissão, pois falta (1) decidir quais e como expor os dados aos leitores, (2) fazer uma sincronização final com os outros autores, e (3) alguns refinamentos finais como a seção de agradecimentos, polimento do texto, ajuste do tamanho entre outros. Desta forma, entendemos que o trabalho despendido e estimulado nas atividades da disciplina, colocaram no papel dados (coletados há um certo tempo) para a produção de uma base sólida e próxima para a submissão de um artigo científico em uma *venue* importante que é o ICSE 2026.