



Rodrigo Brito de Freitas Lima

**Modernização de Sistemas Legados: Uma
abordagem baseada em análise de logs,
questionários e entrevistas**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática da PUC-Rio.

Orientador: Profa. Simone Diniz Junqueira Barbosa

Rio de Janeiro
Dezembro de 2024



Rodrigo Brito de Freitas Lima

**Modernização de Sistemas Legados: Uma
abordagem baseada em análise de logs,
questionários e entrevistas**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo:

Profa. Simone Diniz Junqueira Barbosa

Orientador

Departamento de Informática – PUC-Rio

Profa. Isabela Gasparini

Universidade do Estado de Santa Catarina – UDESC

Profa. Greis Francly Mireya Silva Calpa

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Rio de Janeiro, 18 de Dezembro de 2024

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, do autor e do orientador.

Rodrigo Brito de Freitas Lima

Graduou-se em informática pela Universidade Pontifícia Universidade Católica (Rio de Janeiro). Trabalhou como programador na Sul América Seguros, responsável pelos sistemas de emissão de seguros massificados, trabalhou como analista de sistemas na EDS(Electronic Data Systems, em diversos projetos e clientes com foco maior na área de Telecom, trabalhou como analista de sistemas na Azul seguros, responsável pelo sistema de sinistro e trabalha na iniciativa ExACTa da PUC-Rio, como coordenador da esteira de Logística da Petrobras.

Ficha Catalográfica

Lima, Rodrigo Brito de Freitas

Modernização de Sistemas Legados: Uma abordagem baseada em análise de logs, questionários e entrevistas / Rodrigo Brito de Freitas Lima; orientador: Simone Diniz Junqueira Barbosa. – 2024.

120 120 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2024.

Inclui bibliografia

1. keywordpre – Teses. 2. keywordpre – Teses. 3. Modernização. 4. Sistemas Legados. 5. Wrapping. 6. Análise de Logs. I. Barbosa, Simone Diniz Junqueira. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

À minha família, pela paciência enquanto estive um pouco afastado alterando o sistema, preparando os documentos e estudando para as provas, em especial para a minha esposa que foi a maior incentivadora da minha entrada no mestrado e um agradecimento especial ao meu pai, Ricardo de Freitas Lima (*in memoriam*), que foi quem começou e depois de criado manteve os dados, com as suas revisões intermináveis, o objeto que tornou esse estudo viável.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Resumo

Lima, Rodrigo Brito de Freitas; Barbosa, Simone Diniz Junqueira. **Modernização de Sistemas Legados: Uma abordagem baseada em análise de logs, questionários e entrevistas**. Rio de Janeiro, 2024. 120p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Sistemas legados ainda têm muita importância para corporações e usuários. O desafio de como modernizá-los é bastante explorado, seja na reescrita total do sistema ou na substituição das tecnologias subjacentes. Tentativas *ad hoc* de modernização de sistemas podem ser caóticas e custosas. Muitos artigos vêm buscando formas de enfrentar esse desafio, mas não encontramos estudos satisfatórios que propusessem uma forma de manter as funcionalidades e construir um novo design de interface e melhor experiência do usuário transformando sistemas estáticos em sistemas inteligentes. Essa dissertação propõe uma abordagem para enfrentar esse desafio, trazer recomendações e tornar os sistemas legados em sistemas inteligentes com baixo custo e baixo esforço, mantendo as suas funcionalidades principais do sistema legado, através de análises de *logs* de uso, questionários e entrevistas com os usuários, identificamos dificuldades, pontos fracos do sistema, comportamento de uso e novas oportunidades de melhoria.

Palavras-chave

Modernização; Sistemas Legados; Wrapping; Análise de Logs.

Abstract

Lima, Rodrigo Brito de Freitas; Barbosa, Simone Diniz Junqueira (Advisor). **Modernization of Legacy Systems: An approach based on log analysis, questionnaires and interviews**. Rio de Janeiro, 2024. 120p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Legacy systems still have a lot of importance for corporations and users. The challenge of how to modernize them is widely explored, whether by completely rewriting the system or replacing the underlying technologies. *ad hoc* attempts to modernize systems can be chaotic and costly. Many articles have been looking for ways to face this challenge, but we have not found satisfactory studies that propose a way to maintain functionalities and build a new interface design and better user experience by transforming static systems into intelligent systems. This dissertation proposes an approach to face this challenge, bring recommendations and turn legacy systems into intelligent systems with low cost and low effort, maintaining the main functionalities of the legacy system, through analysis of *logs* of use, questionnaires and interviews with users, we identify difficulties, system weaknesses, usage behavior and new opportunities for improvement.

Keywords

Modernization; Legacy systems; Wrapping; Log analysis.

Sumário

1	Introdução	15
1.1	Motivação	15
1.2	Definição do Problema	15
1.3	Metodologia	16
1.4	Contribuições	17
1.5	Organização deste Documento	17
2	Fundamentação	19
2.1	Sistemas Legados	19
2.2	Modernização de Sistemas Legados	21
2.2.1	Análise de Logs	23
2.2.2	Regras de Negócio	25
2.3	<i>Wrapping</i>	26
2.3.1	<i>Wrapper</i> de APIs	27
2.3.2	Encapsulamento de Funcionalidades	27
2.3.3	Adaptação de Sistemas Legados	28
2.4	Sistemas de Recomendação	29
3	Trabalhos Relacionados	32
3.1	Procedimento	32
3.2	Execução	33
3.2.1	SQ1 Como realizar uma modernização de sistemas legados sem alteração das regras de negócio?	34
3.2.2	SQ2 Como realizar uma análise de <i>logs</i> eficiente para extrair informações importantes para viabilizar a modernização do sistema?	35
3.3	Considerações sobre os Trabalhos Relacionados	36
4	Abordagem proposta para a Modernização de Sistemas Legados	38
4.1	Introdução	38
4.2	Fase 1 - Coleta de dados	39
4.3	Fase 2 - Questionários e Entrevistas	42
4.4	Fase 3 - Recomendação de Ajustes	43
4.5	Fase 4 - Desenvolvimento e Avaliação	44
4.6	Sistema de análise de logs	45
4.6.1	Requisitos Funcionais do Sistema de Análise de <i>logs</i>	47
4.6.2	Requisitos Não Funcionais do Sistema de Análise de <i>logs</i>	48
4.7	Cenário de Aplicação da Abordagem Proposta	48
4.7.1	Cenário de Sucesso 1	48
4.7.1.1	Situação do Sistema Legado:	48
4.7.1.2	Desenvolvimento da Análise	49
4.7.1.3	Resultados	49
4.7.2	Cenário de Sucesso 2	49
4.7.2.1	Situação do Sistema Legado:	49

4.7.2.2	Resultados	50
4.7.3	Cenário de Falha 1	50
4.7.3.1	Situação do Sistema Legado:	50
4.7.3.2	Desenvolvimento da Análise	51
4.7.3.3	Resultados	51
4.7.4	Cenário de falha 2	51
4.7.4.1	Situação do Sistema Legado	51
4.7.4.2	Desenvolvimento da Análise	52
4.7.4.3	Resultados	52
4.8	Considerações	52
5	Estudo de aplicação da abordagem proposta	53
5.1	Objeto de estudo	53
5.2	Planejamento do Estudo Preliminar	54
5.2.1	Objetivo	55
5.2.2	Procedimento	55
5.2.2.1	Diagnóstico Inicial	55
5.2.2.2	Coleta e análise dos dados de <i>logs</i> , questionários e entrevistas	56
5.2.2.3	Elaboração das recomendações de modernização	57
5.3	Resultados do Estudo Preliminar	58
5.3.1	Análise de <i>logs</i>	58
5.3.2	Questionário	62
5.4	Sugestão de Modernização	64
5.4.1	Sistema de recomendação por conteúdo	64
5.4.2	Criação de canais de comunicação via rede social	65
5.4.3	Mudança na forma de consulta dos atributos primários	65
5.4.4	Melhoria da identidade visual do site	66
5.5	Avaliação das Modernizações Realizadas	67
5.5.1	Avaliação do Sistema de Recomendação	67
5.5.1.1	Objetivo	67
5.5.1.2	Procedimento	68
5.5.1.3	Material	68
5.5.1.4	Fonte de Dados	69
5.5.1.5	Resultados	69
5.5.1.6	Execução	71
5.5.1.7	Análise dos Dados Quantitativos	73
5.5.2	Avaliação do Redesign das Consultas	73
5.5.2.1	Objetivo	73
5.5.2.2	Procedimento	73
5.5.2.3	Material	75
5.5.2.4	Fonte de Dados	75
5.5.2.5	Resultados	75
5.5.2.6	Execução	76
5.5.2.7	Análise dos Dados Quantitativos	77
5.5.3	Avaliação da Nova Identidade Visual do Site	79
5.5.3.1	Objetivo	79
5.5.3.2	Procedimento	79
5.5.3.3	Material	80
5.5.3.4	Tela principal de visualização do jogo - Resultado	80

5.5.3.5	Organização da seção dos Últimos 12 Jogos - Resultados	81
5.5.3.6	Telas de apresentação dos atributos primários - Resultados	82
5.5.3.7	Forma de acesso à seção de especiais - Resultados	83
5.5.3.8	Nova visualização de rankings - Resultados	84
5.5.3.9	Nova visualização de consultas combinadas – Resultados	85
5.5.4	Participantes	85
5.5.5	Execução	86
5.5.6	Análise dos Dados Quantitativos	88
5.5.7	Discussão	89
5.5.8	Avaliação dos Canais de Comunicação do Site	89
5.5.8.1	Objetivo	92
5.5.8.2	Procedimento	93
5.5.8.3	Material	94
5.5.8.4	Participantes	94
5.5.8.5	Execução	94
5.5.8.6	Análise dos Dados Quantitativos	94
5.5.8.7	Resultados	95
5.5.8.8	Discussão	95
5.5.9	Considerações sobre as Avaliações	96
6	Avaliação da Abordagem por Analistas Externos	97
6.1	Objetivo	97
6.2	Participantes	97
6.3	Material	97
6.4	Procedimento	97
6.5	Discussão dos resultados	98
6.5.1	Participante 1 - Aluno do primeiro período	98
6.5.2	Participantes 2 e 3 - Profissionais com larga experiência na análise de sistemas e criação e manutenção de sistemas, especialmente em Cobol.	100
6.5.3	Resumo	102
6.5.4	Respostas às Questões de Pesquisa	102
7	Conclusões	104
7.1	Questões de Pesquisa	104
7.2	Contribuições	105
7.2.1	Abordagem de Modernização	105
7.2.2	Sistema para Análise de <i>logs</i>	105
7.3	Trabalhos Futuros	106
	Referências Bibliográficas	106
A	Material do Estudo	112
A.1	Termo de Consentimento Livre e Esclarecido	112
A.2	Questionário para Captura de Opiniões sobre o Sistema	112
A.3	Arquivos de Logs coletados e analisados	112
A.4	Documento de Sugestão	113
A.5	Documento de Sugestão Formulário	113
A.6	Questionário para Captura de Opiniões sobre o Sistema Alterado	113

A.7	Texto da Abordagem para avaliação	113
A.8	Roteiro de entrevista	113
A.9	Google colab usado nas avaliações de abordagem	113
B	Códigos Utilizados neste Trabalho	115
B.1	Informações sobre o código	115
B.1.1	Gera referência	115
B.1.2	Análise de títulos por dia	115
B.1.3	Análise de títulos por hora	116
B.1.4	Análise de acesso a páginas por dia	117
B.1.5	Análise dos acessos ao sistema	117
B.1.6	Análise de Títulos com pouco ou nenhum acesso	117
B.1.7	Análise da tecnologia utilizada para acessar o sistema	118
B.1.8	Identificação de links - Documentação	118
B.1.9	Análise de resposta aos questionários	118
B.1.10	Análise de retorno das postagens em redes sociais	119
B.1.11	Análise de atividades realizadas no sistema	119
B.1.12	Análise de acessos ao sistema de recomendação	119
B.1.13	Análise de acessos as consultas primárias do sistema	120

Lista de Figuras

Figura 4.1	Abordagem de modernização proposta	38
Figura 4.2	Abordagem de modernização - Fase 1: Coleta de dados	40
Figura 4.3	Abordagem de modernização – Fase 2	42
Figura 4.4	Abordagem de modernização – Fase 3 - Recomendação de Ajustes	44
Figura 4.5	Abordagem de modernização – Fase 4 - Desenvolvimento e avaliação de ajustes	44
Figura 5.1	Histórico do site <code>fluzao.xyz</code>	53
Figura 5.2	Desafios de Modernização	56
Figura 5.3	Acessos às páginas de Inter e Palmeiras	60
Figura 5.4	Acessos totais aos principais adversários	60
Figura 5.5	Acessos totais ao árbitro Wilmar Roldan	61
Figura 5.6	Acessos totais ao goleiro Paulo Goulart	61
Figura 5.7	Acessos totais ao Bayern Munique	62
Figura 5.8	Tecnologia usada para acessar site	63
Figura 5.9	Avaliação sobre implementar sistema de recomendação	63
Figura 5.10	Engajamento médio diário	69
Figura 5.11	Engajamento médio mensal	70
Figura 5.12	Preferência dos respondentes sobre a tela principal	70
Figura 5.13	Avaliação da tela principal	71
Figura 5.14	Avaliação consultas relevantes na tela principal	72
Figura 5.15	Recomendações feitas pelo sistema de recomendação primário	72
Figura 5.16	Recomendações feitas pelo sistema de recomendação Secundário	73
Figura 5.17	Número total de cliques por dia	76
Figura 5.18	Número total de cliques por mês	76
Figura 5.19	Preferência entre telas de consulta	77
Figura 5.20	Avaliação da nova tela de consulta	77
Figura 5.21	Nova tela de consulta	78
Figura 5.22	Preferência sobre a tela de jogo	80
Figura 5.23	Avaliação da nova tela de jogo	81
Figura 5.24	Tela de jogo antiga	81
Figura 5.25	Tela de jogo nova	82
Figura 5.26	Preferência do novo formato dos últimos 12 jogos	82
Figura 5.27	Avaliação do novo formato dos últimos 12 jogos	83
Figura 5.28	Novo formato dos últimos 12 jogos	83
Figura 5.29	Preferência da apresentação dos atributos primários	84
Figura 5.30	Avaliação da nova apresentação dos atributos primários	84
Figura 5.31	Nova tela de apresentação dos adversários	85
Figura 5.32	Novo formato de acesso à seção de especiais	86
Figura 5.33	Seção de especiais antiga	86
Figura 5.34	Preferência da apresentação da seção especiais	87
Figura 5.35	Avaliação da nova apresentação da seção especiais	87

Figura 5.36	Avaliação da distribuição das consultas da seção especiais	87
Figura 5.37	Nova distribuição da seção Especiais	88
Figura 5.38	Novo ranking de jogadores	89
Figura 5.39	Novo ranking de adversários	90
Figura 5.40	Preferência da apresentação de ranking de jogadores	90
Figura 5.41	Avaliação da nova apresentação de ranking de jogadores	91
Figura 5.42	Preferência da apresentação de ranking de adversários	91
Figura 5.43	Avaliação da nova apresentação de ranking de adversários	91
Figura 5.44	Nova tela de visualização de especiais	92
Figura 5.45	Preferência da apresentação da visualização de especiais	92
Figura 5.46	Avaliação da nova apresentação da visualização de especiais	93
Figura 5.47	Visualização postagem rede social X	95
Figura 5.48	Avaliação do conteúdo postado na rede social X	96

Lista de Tabelas

Tabela 3.1	Principais características dos trabalhos relacionados.	37
------------	--	----

1

Introdução

Os sistemas legados são sistemas complexos e difíceis de manter devido à escassez de habilidades e à arquitetura de código com camadas fortemente acopladas, que são uma indicação da obsolescência da tecnologia do sistema. Esses sistemas ficam em uso por muitos anos e foram desenvolvidos para ter uma longa vida útil. Em virtude de muitos avanços da tecnologia, essa infraestrutura não atende às necessidades dos usuários. De acordo com o engenheiro de sistemas Feathers (2004), código legado é um código sem testes automatizados, e os principais desafios dos processos legados são débito técnico e trabalho manual, ou seja, eles dependem excessivamente da intervenção humana e deixam a desejar nas suas configurações.

1.1

Motivação

A motivação desta pesquisa é apresentar uma abordagem de modernização de sistemas legados que transforme a experiência do usuário sem alterar as regras de negócio com baixo custo e esforço.

A modernização de sistemas legados se concentra em realizar uma completa reformulação dos sistemas, utilizando de diversas estratégias e ferramentas, como engenharia reversa para transformar o sistema legado em uma arquitetura de microsserviços ou Arquitetura orientada a serviços (SOA) e ferramentas para a extração de regras de negócio. Devido à complexidade dos programas legados, regras de negócio importantes podem ser quebradas ou ignoradas, tornando a experiência do usuário com o novo sistema desagradável e podendo fazer com que todo o esforço tenha sido em vão, uma vez que a nova ferramenta pode ser abandonada por falta de confiança ou usabilidade.

1.2

Definição do Problema

Sistemas legados são plataformas ou softwares que estão obsoletos, porém ainda possuem um grande valor para as organizações ou usuários. São sistemas que produzem valor mas, por serem estáticos e engessados, não trazem a experiência que os usuários esperam nos dias de hoje. Esses sistemas costumam

apresentar programas complexos que contemplam as principais regras de negócios. Como são escritos em linguagens de programação mais antigas e que nem sempre são muito conhecidas pelos desenvolvedores atuais, tornam as tarefas de modernização difíceis de conduzir e muito custosas, principalmente devido ao retrabalho gerado quando se apresenta uma nova proposta de melhoria da experiência do usuário, proposta esta que nem sempre está diretamente ligada às funcionalidades de negócio do sistema original.

Pensando em todos esses problemas, apresentamos nesta dissertação uma nova abordagem de modernização visando melhorar a experiência do usuário (UX). Trata-se de uma abordagem de análise de sistemas legados através de análise de *logs*, questionários e entrevistas com usuários, buscando informações de navegação pelos sistemas para entender como eles são utilizados, quais as preferências de acesso dos usuários, quais as dificuldades encontradas na navegação, que áreas do sistema não são acessadas e por quê. Essa análise nos permite então combinar todo esse material e sugerir modernizações no front end, trazendo uma nova experiência do usuário, transformando sistemas legados em uma interface mais interativa, mantendo as funcionalidades principais do sistema, com baixo custo e esforço.

Definimos então as seguintes questões de pesquisa:

RQ1: Como modernizar sistemas legados através de uma análise de logs, questionários e entrevistas?

RQ2: Como a alteração de uma interface através de sugestões levantadas pela análise de *logs*, questionários e entrevistas pode contribuir para uma melhor experiência do usuário?

1.3

Metodologia

Foram realizadas pesquisas bibliográficas através de sites de busca de artigos acadêmicos utilizando as palavras-chave Modernization, Legacy Systems, Log Analysis e Wrapping, já aplicando os critérios de inclusão e exclusão definidos na Seção 3.2.

Foi feita uma análise sobre as ferramentas de análise de logs no sentido de buscar uma ferramenta adequada com custo zero, identificamos o Google Analytics ¹ e o configuramos no sistema escolhido para a pesquisa ². Durante as três semanas iniciais realizamos diversos estudos com o Google Analytics, identificando relatórios já existentes e criando novos relatórios de interesse. Após três semanas de coleta inicial foi feita a primeira extração de logs.

¹<https://analytics.google.com/>

²<https://www.fluzao.xyz>

Devido à limitação da ferramenta de coleta de logs, que permite realizar o download somente de 5.000 linhas de cada vez, determinamos a periodicidade semanal. Realizamos semanalmente os downloads de logs por um ano, de 03/08/2023 até 02/08/2024.

Criamos programas de análise seguindo as necessidades identificadas nas análises dos arquivos e relatórios gerados no Google Analytics, Elaboramos formulários para questionários e coletamos dados dos usuários do sistema em tempos determinados em 2 períodos, setembro de 2023 e julho de 2024. Com base nos dados coletados, elaboramos um roteiro de entrevista que permitiu avaliar dúvidas que levantamos nas análises dos dados dos questionários. A análise dos dados nos permitiu criar a abordagem de modernização de sistemas legados que é objeto deste trabalho. Avaliamos a abordagem através de estudos realizados nos resultados obtidos pelas modernizações propostas no site ³ e com a avaliação de dois profissionais com experiência em sistemas legados e um aluno de graduação de segundo período.

1.4

Contribuições

A principal contribuição deste trabalho é a definição de uma nova abordagem baseada na análise de logs, questionários e entrevistas com usuários do sistema, que possibilite a elaboração de recomendações de modernização para transformar o sistema legado em um sistema inteligente, tendo como base os dados coletados pelo uso frequente da ferramenta.

Além disso, a abordagem visa a modernizar sistemas legados sem alterar o seu ponto central, que são as regras de negócio, elevando o nível de confiança da nova ferramenta que estará aderente tanto a boas práticas de UX quanto às regras de negócio já estabelecidas com baixo custo e esforço.

1.5

Organização deste Documento

O restante deste documento está organizado da seguinte maneira: O Capítulo 2 apresenta os principais conceitos, que fundamentam este trabalho. O Capítulo 3 apresenta os trabalhos relacionados, se concentrando nos tópicos Modernização, Análise de Logs e Wrapping. Em seguida, o Capítulo 4 descreve a abordagem proposta para realizar a modernização dos sistemas legados. O Capítulo 5 descreve a aplicação da abordagem, desde a análise de logs até a aplicação do questionário, depois no Seção 5.4 listamos as sugestões de redesign e suas origens. Os resultados das primeiras implementações estão no Seção 5.5.

³<https://www.fluzao.xyz>

As avaliações da abordagem estão descritas no Capítulo 6. Finalmente, o Capítulo 7 retoma as principais contribuições deste trabalho e propõe direções para trabalhos futuros.

2

Fundamentação

Este capítulo apresenta os principais conceitos que fundamentam nosso trabalho: sistemas legados (Seção 2.1); modernização de sistemas legados (Seção 2.2); *wrapping* (Seção 2.3) e sistemas de recomendação (Seção 2.4).

2.1

Sistemas Legados

Os sistemas legados referem-se a tecnologias e plataformas de software que foram desenvolvidas há vários anos e que, apesar de sua idade, ainda são utilizadas nas operações diárias de muitas organizações (Althani e Khaddaj, 2017; Bakar et al., 2019). Esses sistemas, muitas vezes, são baseados em tecnologias que foram superadas por inovações mais recentes. Como permanecem em operação devido ao seu papel crítico e à complexidade envolvida em sua substituição, é fundamental entender e gerenciar esses sistemas para a continuidade dos negócios e para a adaptação às novas demandas tecnológicas.

Esta dissertação explora as características dos sistemas legados, seus impactos na eficiência organizacional e as estratégias para sua modernização. A análise considera aspectos técnicos, econômicos e operacionais, oferecendo uma visão abrangente dos desafios e soluções associados a esses sistemas (Feathers, 2004).

Os sistemas legados são frequentemente caracterizados por:

- Tecnologia Obsoleta: Muitas vezes, esses sistemas utilizam hardware e software desatualizados que não são mais suportados pelos fabricantes (Abdellatif et al., 2021; Althani e Khaddaj, 2017; Bakar et al., 2019; Bakar e Razali, 2013; Leon e Horita, 2021; Strobl et al., 2020; De Vargas Agilar et al., 2016).
- Complexidade: Eles frequentemente foram desenvolvidos em uma época em que os padrões de desenvolvimento e manutenção eram menos rigorosos, resultando em código complicado e difícil de entender (Klammer e Pichler, 2014; Hussain et al., 2017; Baghdadi e Al-Bulushi, 2015; Mateos et al., 2019; Pereira et al., 2023; Hussain et al., 2017; Kazanavičius e Mažeika, 2023; Baghdadi e Al-Bulushi, 2015; Jomhari et al., 2024; Hus-

sain et al., 2017; Somogyi e Kovesdan, 2021; Kazanavičius e Mažeika, 2023; Medeiros et al., 2023; Trabelsi et al., 2023).

- Integração Restrita: A integração com tecnologias modernas pode ser limitada, tornando a comunicação com novos sistemas um desafio (Abdellatif et al., 2021; Althani e Khaddaj, 2017; Bakar e Razali, 2013; Leon e Horita, 2021; Strobl et al., 2020; De Vargas Agilar et al., 2016; Ponnusamy e Eswararaj, 2023).
- Dependência Crítica: Apesar de serem antigos, esses sistemas muitas vezes são essenciais para as operações diárias e processos de negócio (Althani e Khaddaj, 2017; Bakar e Razali, 2013; Leon e Horita, 2021; Strobl et al., 2020; De Vargas Agilar et al., 2016; Ponnusamy e Eswararaj, 2023).
- Código Complexo e Mal Documentado: Código frequentemente complexo, difícil de entender e com pouca documentação (Klammer e Pichler, 2014; Hussain et al., 2017; Baghdadi e Al-Bulushi, 2015; Mateos et al., 2019; Pereira et al., 2023).

A manutenção de sistemas legados pode ter impactos significativos na eficiência organizacional (Abdellatif et al., 2021; Althani e Khaddaj, 2017; Bakar et al., 2019; Bakar e Razali, 2013; Leon e Horita, 2021; Strobl et al., 2020; De Vargas Agilar et al., 2016):

- Custo de Manutenção: A manutenção desses sistemas pode ser cara devido à falta de suporte e à necessidade de habilidades especializadas.
- Risco de Falhas: A obsolescência pode aumentar a probabilidade de falhas e problemas técnicos que afetam a operação.
- Desafios de Integração: A dificuldade em integrar sistemas legados com tecnologias modernas pode limitar a capacidade da organização de adotar novas soluções e inovar.
- Impacto na Produtividade: Funcionários podem enfrentar dificuldades ao trabalhar com sistemas antigos, o que pode reduzir a produtividade e aumentar a frustração.
- Sistemas que possuem um papel importante e que certamente ajudam na produtividade da empresa e sua falta pode causar grandes impactos.

A gestão de sistemas legados é um desafio complexo que exige uma abordagem cuidadosa e estratégica. Apesar dos problemas associados a esses sistemas, eles continuam desempenhando um papel vital em muitas organizações. A modernização dos sistemas legados deve ser planejada de forma a equilibrar

os custos, riscos e benefícios, garantindo que a transição para novas tecnologias seja realizada de maneira eficaz e eficiente. Desta forma, os sistemas legados são o principal objeto de estudo desta pesquisa, que propõe e avalia uma abordagem que suporta o processo de modernização dos sistemas com baixo custo e esforço, mantendo as suas funcionalidades para que os sistemas possam continuar desempenhando o papel importante que têm para as organizações e usuários.

2.2

Modernização de Sistemas Legados

Sistemas legados desempenham um papel crucial em muitas organizações, porém enfrentam uma série de desafios devido à sua obsolescência tecnológica e arquitetônica. Nesta dissertação, discutimos estratégias de refatoração, reengenharia e substituição, além das implicações econômicas e operacionais de cada abordagem. A análise é sustentada por estudos de caso e melhores práticas do setor.

A modernização dos sistemas legados é um campo de estudo importante devido ao impacto que esses sistemas têm na eficiência operacional e na capacidade de inovação das empresas.

A gestão e a modernização de sistemas legados envolvem vários desafios, incluindo (Abdellatif et al., 2021; Althani e Khaddaj, 2017; Bakar e Razali, 2013; Leon e Horita, 2021; De Vargas Agilar et al., 2016):

- **Manutenção e Suporte:** A manutenção contínua de sistemas legados pode ser cara e exigir habilidades especializadas que se tornam cada vez mais raras (Bakar et al., 2019; Strobl et al., 2020).
- **Integração e Compatibilidade:** Dificuldades em integrar sistemas legados com novas soluções tecnológicas podem limitar a capacidade de inovação (Strobl et al., 2020; Ponnusamy e Eswararaj, 2023).
- **Riscos de Segurança:** Sistemas legados podem ter vulnerabilidades de segurança que não são facilmente corrigidas devido à obsolescência das tecnologias utilizadas.
- **Custo:** O custo de operação e manutenção de sistemas legados pode ser alto, especialmente quando comparado com soluções mais modernas e eficientes.

Diversas estratégias podem ser empregadas para a modernização de sistemas legados, cada uma com seus próprios benefícios e desafios. Existem várias abordagens para lidar com sistemas legados e sua modernização

- Reengenharia: Consiste na reescrita ou reconstrução parcial do sistema, preservando a funcionalidade essencial, mas modernizando a arquitetura e o código. A reengenharia pode melhorar significativamente a performance e a escalabilidade, mas pode exigir um investimento considerável de tempo e recursos (Bowen, 2015; García-Borgoñón et al., 2023; Brown, 1993; Bianchi et al., 2003; Visaggio, 2001; Doderó et al., 2019; Medeiros et al., 2023; Trabelsi et al., 2023).
- Refatoração: Envolve a reestruturação do código existente para melhorar a legibilidade e a manutenção sem alterar a funcionalidade externa. Beneficia a organização ao tornar o código mais gerenciável, mas pode não resolver todos os problemas associados ao sistema legado (Pereira et al., 2023; Hussain et al., 2017; Kazanavičius e Mažeika, 2023; Baghdadi e Al-Bulushi, 2015).
- Substituição Gradual: Algumas organizações optam por uma abordagem híbrida, migrando gradualmente para novas soluções enquanto mantêm o sistema legado funcionando. Isso pode reduzir os riscos e os custos iniciais, mas pode ser complexo de gerenciar e coordenar (Klammer e Pichler, 2014; Hussain et al., 2017; Baghdadi e Al-Bulushi, 2015; Mateos et al., 2019).
- Substituição Total: Envolve a substituição completa do sistema legado por uma nova solução. Embora possa ser a abordagem mais dispendiosa e arriscada, permite à organização implementar tecnologias e práticas modernas desde o início (Jomhari et al., 2024; Hussain et al., 2017; Somogyi e Kovesdan, 2021; Kazanavičius e Mažeika, 2023; Medeiros et al., 2023; Trabelsi et al., 2023).
- Virtualização: Uso de tecnologias de virtualização para preservar a funcionalidade dos sistemas legados em um ambiente mais moderno (Hussain et al., 2017; Visaggio, 2001; Somogyi e Kovesdan, 2021; Medeiros et al., 2023).
- Integração de Sistemas: Desenvolvimento de interfaces e adaptadores que permitam a comunicação entre sistemas legados e novas tecnologias (Hussain et al., 2017; Visaggio, 2001; Baghdadi e Al-Bulushi, 2015; Mateos et al., 2019).

Sistemas legados são uma parte crítica da infraestrutura tecnológica de muitas organizações, mas sua manutenção e modernização são desafios complexos. Propomos neste trabalho uma abordagem estratégica e planejada para a modernização com a intenção de ajudar as organizações a superar esses desafios e a melhorar sua eficiência operacional e capacidade de inovação. Nossa

abordagem de modernização de sistemas legados tem com base análise de logs de navegação do sistema (Subseção 2.2.1), aplicação de questionários e condução de entrevistas com os usuários do sistema. Permite gerar recomendações de modernização sem alteração das regras de negócio (Subseção 2.2.2), com baixo custo e esforço utilizando conceitos da técnica de *wrapping* (Seção 2.3).

2.2.1 Análise de Logs

A análise de logs é o processo de examinar e interpretar registros gerados por sistemas de computação e redes para obter informações úteis para a administração, monitoramento e segurança (Visaggio, 2001). Logs, que podem incluir eventos de sistema, transações de usuário e erros, apoiam a compreensão do funcionamento interno dos sistemas e para a detecção precoce de problemas e ameaças.

A análise de logs é uma prática importante na administração de sistemas de informação, oferecendo insights valiosos sobre o comportamento dos sistemas e a segurança (Breitmayer et al., 2023; Apaolaza e Vigo, 2017).

Os logs são classificados em diferentes categorias, cada uma com seus próprios formatos e finalidades. Entre os principais tipos estão (Allam, 2018; Jayathilake, 2012; Miransky et al., 2016):

- Logs de Sistema: Registram eventos relacionados ao funcionamento do sistema operacional, como inicializações, encerramentos e falhas.
- Logs de Aplicação: Capturam informações sobre a execução de aplicações, incluindo mensagens de erro, transações e operações executadas.
- Logs de Segurança: Envolvem registros relacionados a segurança, como autenticações, acessos e tentativas de intrusão.
- Logs de Rede: Contêm informações sobre o tráfego de rede, incluindo pacotes enviados e recebidos, e possíveis problemas de conectividade.

A análise de logs pode ser realizada por meio de diversos métodos, cada um adequado para diferentes objetivos (Allam, 2018; Jayathilake, 2012; Miransky et al., 2016):

- Filtragem e Busca: Tratam-se de técnicas básicas que envolvem a seleção e visualização de registros relevantes a partir de grandes volumes de dados. Ferramentas de busca e filtragem ajudam a localizar eventos específicos e padrões de interesse.
- Análise Estatística: Utiliza métodos estatísticos para identificar padrões e anomalias nos logs. Técnicas como a análise de séries temporais podem ser usadas para detectar variações e tendências.

- Correlação de Eventos: Envolve a combinação de diferentes fontes de logs para identificar relações e dependências entre eventos. Isso pode ajudar a formar uma visão mais completa e integrada dos problemas e incidentes.
- Análise de Anomalias: Emprega algoritmos para identificar comportamentos incomuns que podem indicar problemas ou ataques. Técnicas de aprendizado de máquina, como *clustering* e detecção de *outliers*, são comuns nessa abordagem.

Na abordagem que propomos, utilizamos as três primeiras técnicas listadas acima, isolando eventos relevantes dentre uma grande quantidade de registros e identificando, através de análise estatística, os padrões de navegação dos sistemas legados. Os métodos foram escolhidos devido à natureza da abordagem e o foco é na análise de dos padrões de navegação. Desta forma filtramos os melhores logs, correlacionamos informações e analisamos os dados estatísticos correspondentes. Neste trabalho não demos foco em possíveis anomalias.

A análise de logs tem diversas aplicações práticas em ambientes de TI e segurança (Allam, 2018; Jayathilake, 2012; Miransky et al., 2016):

- Diagnóstico de Problemas: Identificação e solução de problemas técnicos e falhas no sistema por meio da revisão de logs e análise de eventos registrados.
- Monitoramento de Performance: Avaliação do desempenho do sistema e detecção de degradações ou gargalos através da análise contínua de logs.
- Segurança e *Compliance*: Detecção de atividades suspeitas e conformidade com políticas de segurança e regulamentos. Logs de segurança são essenciais para a auditoria e resposta a incidentes.
- Otimização de Recursos: Análise de padrões de uso e desempenho para otimizar a alocação de recursos e melhorar a eficiência operacional.
- Identificação de Padrões de Navegação, indicando os principais caminhos percorridos no sistema, visualizando as maiores dificuldades e viabilizando análises para indicar os melhores rumos de modernização dos sistemas.

A análise de logs enfrenta diversos desafios:

- Volume de Dados: O grande volume de logs pode tornar a análise complexa e lenta. Estratégias de agregação e amostragem podem ser necessárias para lidar com grandes volumes de dados (Miransky et al., 2016).

- Variedade e Complexidade: Logs podem variar em formato e conteúdo, tornando a integração e análise desafiadoras. A normalização e padronização dos dados são práticas recomendadas (Allam, 2018; Miransky et al., 2016).
- Privacidade e Segurança: A análise de logs deve respeitar as regulamentações de privacidade e proteger informações sensíveis. É importante implementar controles de acesso e criptografia (Allam, 2018; Miransky et al., 2016).

Algumas formas para se realizar uma boa coleta de logs envolvem a instrumentação do próprio sistema para a captura de alguns dados importantes de sua navegação ou a implantação de ferramentas e scripts para automação, reduzindo a carga de trabalho manual e melhorando a eficiência. No entanto, para que isso seja possível, o sistema precisa ter uma documentação detalhada e robusta.

A análise de logs é uma prática importante para a gestão e segurança de sistemas de informação, oferecendo insights cruciais para a resolução de problemas e a melhoria contínua. A adoção de métodos apropriados e ferramentas eficazes pode maximizar os benefícios da análise de logs e ajudar as organizações a manter a integridade e o desempenho de seus sistemas.

A análise de logs é a principal ferramenta utilizada pela abordagem com papel essencial na identificação de padrões de navegação que vão guiar as sugestões de modernização dos sistemas conforme detalhado na seção Capítulo 4.

2.2.2

Regras de Negócio

Regras de negócio são diretrizes que governam o comportamento e as decisões dentro de um sistema ou processo de negócio. Esta dissertação explora o conceito de regras de negócio, sua importância na modelagem de sistemas e processos, e as melhores práticas para sua implementação e gerenciamento. A análise abrange desde a definição e categorização das regras até as abordagens para sua integração em sistemas de software, com o objetivo de garantir que os sistemas de informação estejam alinhados com as necessidades e objetivos de negócios.

Regras de negócio podem ser definidas como as condições e ações que governam o funcionamento dos processos de negócio. Elas podem ser classificadas em diferentes tipos, incluindo (Herbst et al., 1994; Taveter e Wagner, 2001):

- Regras de Validação: Definem condições que os dados devem atender para serem considerados válidos. Exemplo: uma regra pode definir que o cadastro só será efetivado se o CPF estiver indicado.
- Regras de Derivação: Determinam como derivar novos dados a partir dos dados existentes. Exemplo: uma regra pode definir que o imposto sobre vendas é calculado como uma porcentagem do preço total do produto
- Regras de Processamento: Especificam como os dados devem ser processados ou manipulados durante a execução de um processo de negócios. Exemplo: uma regra vai indicar que, se o cliente escolhe o plano de pagamento X, vai seguir o processamento do plano X.
- Regras de Decisão: Orientam o fluxo de decisões e ações com base em condições específicas. Exemplo: uma regra pode indicar que, se um cliente não pagar uma fatura dentro de 30 dias, a conta deve ser encaminhada para cobrança.

As regras de negócio desempenham um papel crucial na definição e gestão dos processos de uma organização. A implementação eficaz dessas regras garante que os sistemas de informação estejam alinhados com as políticas e objetivos de negócios, promovendo eficiência e consistência. Desta forma, a abordagem que propomos tem como um dos seus pilares a manutenção das regras de negócio da forma como elas existem no sistema, reduzindo assim possíveis focos de insatisfação dos usuários com o sistema modernizado.

2.3

Wrapping

O *wrapping* é uma técnica de engenharia de software. Wrapping envolve a criação de camadas de encapsulamento ao redor de sistemas, APIs ou funcionalidades. Essas camadas podem fornecer uma interface mais adequada, adaptar funcionalidades a novos contextos ou integrar sistemas legados com novas tecnologias. Esta seção apresenta diferentes formas de *wrapping*, suas vantagens e desafios, e boas práticas para sua implementação.

Wrapping refere-se à prática de criar uma camada de abstração ao redor de um componente de software para modificar, adaptar ou melhorar sua funcionalidade. Existem diferentes formas de *wrapping*, cada uma com seus próprios objetivos e técnicas, são elas:

- *Wrapper* de API (Subseção 2.3.1): Envolve a criação de uma camada de interface que se comunica com uma API existente. O *wrapper* simplifica a interação com a API, fornece uma interface mais adequada e pode

adicionar funcionalidades adicionais (De Lucia et al., 2008; Noffsinger et al., 1998; Sneed, 2000, 2009; Sneed e Verhoef, 2019; Althani e Khaddaj, 2017).

- Encapsulamento de Funcionalidades (Subseção 2.3.2): Refere-se à criação de uma camada ao redor de um componente ou módulo para ocultar a complexidade e fornecer uma interface mais simples e coesa para os usuários ou outros componentes do sistema (De Lucia et al., 2008; Noffsinger et al., 1998; Sneed, 2000, 2009; Sneed e Verhoef, 2019; Althani e Khaddaj, 2017).
- Adaptação de Sistemas Legados (Subseção 2.3.3): Consiste em criar camadas de *wrapping* para integrar sistemas legados com novas tecnologias, permitindo que sistemas antigos funcionem com novas aplicações e plataformas (De Lucia et al., 2008; Noffsinger et al., 1998; Sneed, 2000, 2009; Sneed e Verhoef, 2019; Althani e Khaddaj, 2017).

2.3.1

Wrapper de APIs

O *wrapper* de APIs é uma técnica comum usada para facilitar a interação com APIs complexas ou de baixo nível. Os principais objetivos e benefícios incluem:

- Simplicidade: Proporciona uma interface mais simples e intuitiva para usuários e desenvolvedores, escondendo detalhes complexos da API subjacente.
- Extensibilidade: Permite a adição de novas funcionalidades ou modificações sem alterar a API original. Os wrappers podem incluir funcionalidades adicionais, como validação de dados e manipulação de erros.
- Compatibilidade: Facilita a integração de APIs de terceiros com sistemas existentes, garantindo que a comunicação entre diferentes componentes do sistema seja consistente e eficiente.

2.3.2

Encapsulamento de Funcionalidades

O encapsulamento de funcionalidades é uma técnica utilizada para criar uma camada de abstração em torno de módulos ou componentes de software. Os principais benefícios incluem:

- Modularidade: Permite que os componentes sejam desenvolvidos e mantidos de forma independente, promovendo a modularidade e a reutilização do código.

- Facilidade de Manutenção: Facilita a manutenção e atualização dos componentes, pois mudanças na implementação interna não afetam as interfaces públicas.
- Segurança: Ajuda a proteger o sistema contra alterações indesejadas, expondo apenas as funcionalidades necessárias e escondendo a complexidade interna.

2.3.3

Adaptação de Sistemas Legados

A adaptação de sistemas legados através de *wrapping* é crucial para integrar tecnologias antigas com novas soluções. Os principais objetivos e benefícios incluem:

- Integração: Permite que sistemas legados interajam com novas tecnologias e plataformas, garantindo que os investimentos em tecnologia antiga não sejam perdidos.
- Modernização: Facilita a modernização de sistemas legados, permitindo a implementação de novas funcionalidades e a atualização de interfaces sem alterar o código legado.
- Interoperabilidade: Melhora a interoperabilidade entre sistemas antigos e novos, permitindo que diferentes sistemas funcionem em conjunto de maneira mais eficiente.

Apesar de suas vantagens, o *wrapping* enfrenta vários desafios:

- Complexidade Adicional: A criação de camadas de *wrapping* pode adicionar complexidade ao sistema, o que pode afetar o desempenho e a manutenção.
- Desempenho: O uso de *wrappers* pode introduzir sobrecarga adicional, impactando o desempenho do sistema. Técnicas de otimização e análise de impacto são essenciais para mitigar esses problemas.
- Gerenciamento de Mudanças: A manutenção e atualização dos *wrappers* devem ser cuidadosamente gerenciadas para garantir que continuem compatíveis com as APIs subjacentes e os sistemas integrados.

Para uma implementação bem-sucedida de *wrapping* os seguintes requisitos são desejáveis:

- Design Claro e Consistente: Defina uma interface clara e consistente para o *wrapper*, facilitando a utilização e a integração com outros componentes.

- Documentação Adequada: Mantenha uma documentação completa e atualizada para o *wrapper*, incluindo informações sobre sua funcionalidade, interfaces e exemplos de uso.
- Testes Rigorosos: Realize testes rigorosos para garantir que o *wrapper* funcione corretamente em diferentes cenários e integre-se de forma eficiente com os componentes existentes.
- Monitoramento e Manutenção: Implemente práticas de monitoramento e manutenção para garantir que o *wrapper* continue a atender às necessidades do sistema e a evoluir conforme necessário.

O conceito de *wrapping* desempenha um papel importante na engenharia de software, oferecendo soluções para a integração de sistemas, a simplificação de interfaces e a modernização de tecnologias antigas. Compreender as técnicas, benefícios e desafios associados ao *wrapping* é importante para a criação de soluções de software eficientes e sustentáveis.

O conceito de encapsulamento de funcionalidades é mais utilizado nos processos de modernização baseado em criação de microsserviços; desta forma, não é aderente ao foco desta pesquisa, que tem um foco em, tanto quanto possível, não alterar o código legado.

O conceito de *Wrapper* de APIs é aplicado em sistemas que realizam comunicação via APIs com diversos outros sistemas. Com o passar do tempo, essa comunicação se torna complexa, extensa e apresenta problemas de compatibilidade que precisam ser tratados e modernizados. O objeto da nossa pesquisa não realiza comunicação via API com sistemas externos e, dessa forma, essa técnica não é aderente ao nosso trabalho.

A abordagem de modernização que propomos se apoia no conceito adaptação de sistemas legados do *wrapping* para modernização de sistemas legados. O objetivo central da pesquisa é modernizar sem extrair ou alterar regras de negócio, criando uma “casca” de sistema mais moderna capaz de aumentar a experiência do usuário, sem perda de eficiência do sistema e com baixo custo e esforço.

2.4

Sistemas de Recomendação

Os sistemas de recomendação têm se tornado uma ferramenta essencial em diversas aplicações, desde *e-commerce* até plataformas de *streaming* de mídia. Esta seção descreve conceitos dos sistemas de recomendação, abordando suas técnicas principais, como filtragem colaborativa e baseada em conteúdo, e analisando suas aplicações em diferentes contextos. Também discutimos os

desafios enfrentados por esses sistemas, incluindo problemas de escalabilidade e privacidade.

Sistemas de recomendação sugerem produtos, serviços ou conteúdos aos usuários com base em seus interesses e comportamentos passados. Eles desempenham um papel crucial em melhorar a experiência do usuário e aumentar a satisfação e engajamento.

Os sistemas de recomendação buscam prever as preferências dos usuários e fornecer sugestões personalizadas. As principais abordagens são:

- Filtragem Colaborativa: Baseia-se no comportamento de usuários semelhantes. As recomendações são feitas com base nas preferências de outros usuários que compartilham interesses semelhantes. Ela pode ser baseada em usuários ou com itens. Na filtragem colaborativa, o sistema recomenda itens que foram preferidos por usuários semelhantes. Exemplo: se o Usuário A e o Usuário B têm interesses semelhantes, as preferências de A podem sugerir itens para B. Na filtragem colaborativa baseada em item, o sistema recomenda itens semelhantes aos que o usuário já gostou. Exemplo: se o Usuário C gostou de Filme X, o sistema sugere filmes semelhantes (Chaimalas et al., 2023).
- Filtragem Baseada em Conteúdo: Utiliza informações sobre os itens e o perfil do usuário para fazer recomendações. As sugestões são baseadas nas características dos itens e nas preferências explícitas ou implícitas dos usuários. O sistema utiliza as preferências do usuário com base em itens com que ele interage e recomenda itens com características semelhantes. Exemplo: se um usuário gosta de filmes de ficção científica, o sistema sugere outros filmes desse gênero ou analisa atributos dos itens e recomenda itens com características semelhantes às preferidas pelo usuário (Chaimalas et al., 2023; Lubos et al., 2023; Pellegrini et al., 2022; Rashed et al., 2022; Zhu e Van Roy, 2023).
- Modelos Híbridos: Combinam técnicas de filtragem colaborativa e baseada em conteúdo para melhorar a precisão das recomendações e mitigar as limitações de cada abordagem individualmente. Exemplo: um sistema pode combinar a análise de perfil de usuário com preferências de usuários similares, além de realizar uma integração de dados utilizando técnicas de aprendizado de máquina para integrar diferentes tipos de dados e melhorar a qualidade das recomendações (Chaimalas et al., 2023; Lubos et al., 2023; Pellegrini et al., 2022; Rashed et al., 2022; Zhu e Van Roy, 2023).

Os sistemas de recomendações são amplamente aplicados em diversos contextos, incluindo (Buschek et al., 2022; Zhu e Van Roy, 2023; Rashed et al., 2022):

- E-commerce: Recomenda produtos com base no histórico de compras e navegação dos usuários. Exemplo: Amazon sugere produtos relacionados às compras anteriores.
- Streaming de Mídia: Sugere filmes e músicas com base nas preferências e comportamentos dos usuários. Exemplo: Netflix recomenda filmes e séries com base no histórico de visualização.
- Redes Sociais: Sugere amigos, grupos e conteúdos com base nas interações e interesses dos usuários. Exemplo: Facebook sugere amigos com base em conexões e interesses compartilhados.
- Publicidade Online: Personaliza anúncios com base nos interesses e comportamentos dos usuários para aumentar a relevância e a eficácia dos anúncios.
- Sugere conteúdo de informação a ser consumida com base na navegação dos usuários, seja de forma anônima ou personalizada.

Nesta pesquisa, o sistema de recomendação foi uma das sugestões feitas pelo responsável pela modernização do sistema com o suporte da abordagem proposta com base na análise dos *logs*.

O tipo de sistema de recomendação escolhido foi o sistema de recomendação por conteúdo devido à ausência de login. Isso impossibilita identificar os usuários para fazer recomendações por filtragem colaborativa ou híbrida.

3 Trabalhos Relacionados

Este capítulo apresenta os trabalhos encontrados na literatura que se relacionam com este trabalho. A Seção 3.1 apresenta o procedimento adotado para a seleção dos artigos.

3.1 Procedimento

As questões de pesquisa colocadas na Seção 1.2 foram desdobradas nas seguintes subquestões:

SQ1: Como realizar uma modernização de sistemas legados sem alteração das regras de negócio?

SQ2: Como realizar uma análise de logs eficiente para extrair informações importantes para viabilizar o redesign do sistema?

As subseções a seguir discutem as respostas encontradas na literatura para cada sub-questão.

Para selecionarmos os artigos, definimos os seguintes passos:

1. **Definição das condições de busca** incluindo a string de busca e as bibliotecas digitais onde as buscas serão efetuadas.
2. **Definição dos critérios** de inclusão e exclusão (I/E). Esses critérios foram aplicados manualmente, sobre metadados ou conteúdo, avaliando abstract, resumo, introdução e conclusão.
3. **Busca** nas bibliotecas digitais selecionadas.
4. **Remoção de artigos duplicados**. Este passo foi realizado comparando-se os títulos dos artigos.
5. **Leitura e extração dos dados** dos artigos completos.
6. **Discussão** dos resultados.

3.2 Execução

A escolha do tema e posteriormente da busca na literatura se deu da seguinte forma, iniciamos com uma tentativa de utilização do site de estatísticas do esportivas do Fluminense como base para o estudo, verificadas algumas limitações na parte gráfica e de entendimento sobre o que os usuários acessavam no site, colocamos um gerador de logs para buscar alguma informação, em paralelo ao assistir a matéria sobre sistemas inteligentes identificamos em conjunto com as primeiras análises de logs que um sistema de recomendação seria uma boa alternativa de melhoria no sistema, juntando todas essas partes com a minha experiência nas diversas tentativas de modernização de sistemas legados, surgiu a ideia de propor uma nova abordagem que utilizasse a análise de logs em sistemas legados para facilitar a modernização dos sistemas sem que regras de negócio fossem alteradas ou perdidas, tornando o sistema mais moderno e mantendo sua confiabilidade.

Desta forma as palavras chave: "Log Analysis", "Modernization", "Legacy System", surgiram naturalmente como keywords, após indicação do professor Marcos Kalinowski na defesa da proposta incluímos a palavra "Wrapping". Não existe muita literatura sobre o assunto modernização da forma como abordamos portanto escolhemos não conduzir uma revisão formal da literatura e fizemos uma pesquisa nas últimas conferências e escolhemos uma biblioteca para concentrar as buscas.

Dada a questão de pesquisa ("Como modernizar sistemas legados através de uma análise de logs, questionários e entrevistas"), definimos os seguintes termos de busca: Modernization, Log Analysis, Legacy System, Wrapping. Inicialmente, realizamos buscas separadas por cada palavra-chave por entendermos que a natureza de cada uma delas poderia trazer impactos distintos no que estamos pesquisando. Porém, analisando os primeiros resultados identificamos que seria interessante realizar uma nova busca por Modernization e Legacy System combinadas porque, na busca separada, nem sempre conseguimos identificar com clareza os artigos que tratavam de modernização de sistemas legados.

A pesquisa foi realizada na biblioteca digital Scopus.¹ Dada a maior aderência dos artigos relacionados a *wrapping*, realizamos uma rodada de *snowballing* dos artigos encontrados.

Incluímos artigos escritos em inglês e publicados a partir de 2000, que continham pelo menos uma das expressões de interesse (Modernization, Log Analysis, Legacy System e Wrapping) no título, resumo ou palavras-chave.

¹<https://www.scopus.com/search/form.uri?display=basic>

Em seguida, realizamos uma análise detalhada sobre o que é abordado no artigo e sua relevância com a pesquisa. Muitos artigos tratam da criação de ferramentas para extração de regras de negócio, outros abordam uma total reestruturação para tecnologias mais modernas. Em ambos os casos, os artigos têm muito pouca relação com a pesquisa e foram descartados. Outros artigos foram avaliados manualmente, examinando seus títulos, resumos e palavras-chave em busca de pontos que possam agregar na pesquisa.

Adotamos os seguintes critérios de inclusão:

- Artigos cujo título, resumo ou palavras-chave contivesse as expressões de interesse
- Artigos redigidos em inglês
- Artigos publicados a partir de 2000
- Relação com o tema de pesquisa, com base no resumo (*abstract*), introdução e conclusão

Inicialmente encontramos 93 artigos. Após excluirmos os artigos duplicados, ficamos com 80 artigos.

3.2.1

SQ1 Como realizar uma modernização de sistemas legados sem alteração das regras de negócio?

Estudos relacionados com modernização de sistemas legados podem ser divididos em cinco tópicos relacionados, engenharia reversa/extração de regras de negócio, microsserviço, arquitetura SOA, wrapping e análise de logs.

Os artigos associados a revisões de literatura fazem uma busca e trazem as avaliações das estratégias de migração existentes na literatura e suas aplicações na indústria, desde migração total dos sistemas até um simples Wrapping, passando por arquitetura de microsserviços, arquitetura SOA e modelos de avaliação baseado em logs e banco de dados (Abdellatif et al., 2021; Althani e Khaddaj, 2017; Bakar et al., 2019; Bakar e Razali, 2013; Leon e Horita, 2021; Strobl et al., 2020; De Vargas Agilar et al., 2016; Ponnusamy e Eswararaj, 2023).

Para se chegar no objetivo de modernização de sistemas legados, a técnica de engenharia reversa para extração de regras nos sistemas legados surge como o principal pilar para a implantação em qualquer uma das estratégias encontradas, foram propostos vários tipos de engenharia reversa, incluindo a criação de modelos para a extração das regras (Bowen, 2015; García-Borgoñón et al., 2023; Pereira et al., 2023; Klammer e Pichler, 2014; Brown, 1993; Jomhari et al., 2024; Bianchi et al., 2003), passando por identificação de códigos

que possam ser reutilizados combinados com a extração das regras (Hussain et al., 2017), utilização de logs de uso e arquitetura da aplicação (Visaggio, 2001; Dodero et al., 2019), até uso de machine learning para entendimento das regras de negócio (Somogyi e Kovesdan, 2021).

Outras técnicas encontradas são a transformação total dos sistemas legados em sistemas baseados em microsserviços (Kazanavičius e Mažeika, 2023; Medeiros et al., 2023; Trabelsi et al., 2023) ou em arquitetura SOA ((Baghdadi e Al-Bulushi, 2015; Mateos et al., 2019)) fazendo a decomposição total do sistema baseados em ferramentas de extração de regras de negócio.

Todas essas técnicas e estratégias têm como objetivo realizar a modernização de sistemas legados fazendo a troca dos códigos legados por uma programação mais moderna, diferentemente do que estamos propondo. Uma vez que a nossa abordagem preserva o código legado e cria uma interação com o usuário mais moderna, visando assim atingir a criação de um sistema mais integrado com uma melhor experiência do usuário e com baixo custo e esforço no processo de modernização. O que mais se aproxima da nossa visão são os estudos que se baseiam nas técnicas de wrapping, que é a técnica de criar uma casca ao redor do sistema legado podendo manter as regras de negócio intactas (De Lucia et al., 2008; Noffsinger et al., 1998; Sneed, 2000, 2009; Sneed e Verhoef, 2019). Entendemos que a decisão de manter o código legado deve ser uma estratégia de negócio que cada dono do sistema deve realizar. Sistemas escritos em linguagens obsoletas podem não ter mais um custo-benefício de manutenção que justifique manter o código legado, o que naturalmente não levaria a utilização da nossa abordagem. Porém, sistemas que ainda tenham uma vida útil de manutenção seriam o foco de utilização desta abordagem.

3.2.2

SQ2 Como realizar uma análise de logs eficiente para extrair informações importantes para viabilizar a modernização do sistema?

Estudos que contemplem a análise de logs como ferramenta de modernização dos sistemas legados são bastante raros. Temos uma proposta de mineração de processos dentro de arquivos de logs para identificar comportamentos de uso (Breitmayer et al., 2023) e um segundo estudo com base nos logs de utilização do banco de dados da aplicação para identificação do comportamento do usuário (Apaolaza e Vigo, 2017). Ambos os estudos trazem uma abordagem diferente, extrair logs dos sistemas legados através das interações dos usuários com o sistema, banco de dados, para o mesmo problema de realizar uma análise de logs eficiente que possibilite a modernização do sistema legado, a utilização de ambas as abordagens visa trazer respostas sobre como o sistema vem sendo

utilizado, se o comportamento do usuário é o esperado em hipóteses levantadas pelos analistas, porém nenhum deles visa fornecer soluções de modernização do sistema. O nosso estudo traz uma abordagem da análise de logs dos caminhos percorridos pelo usuário no sistema para identificar comportamentos de uso do sistema pelos usuários e, a partir daí, fornecer sugestões de melhoria visando trazer uma melhor experiência no uso do sistema pelo usuário.

3.3

Considerações sobre os Trabalhos Relacionados

A Tabela 3.1 apresenta um quadro comparativo dos trabalhos relacionados. Para efeitos de comparação, incluímos na última linha as características da nossa proposta, conforme descrito no próximo capítulo.

Tabela 3.1: Principais características dos trabalhos relacionados.

Artigo	Engenharia Reversa	Micro-serviço	SOA	Wrapping	Análise de Logs
Modernização de Sistema Legado / Wrapping / Análise de Logs					
Althani e Khaddaj (2017)	X				
Bakar et al. (2019)	X				
Bakar e Razali (2013)	X				
Bowen (2015)		X	X		
García-Borgoñón et al. (2023)		X	X		
Hussain et al. (2017)		X			
Kazanavičius e Mažeika (2023)		X			
Klammer e Pichler (2014)	X				
Medeiros et al. (2023)	X	X			
Pereira et al. (2023)		X			
Somogyi e Kovesdan (2021)	X	X			
Trabelsi et al. (2023)	X	X			
Leon e Horita (2021)	X				
Strobl et al. (2020)	X				
Abdellatif et al. (2021)	X		X		
De Vargas Agilar et al. (2016)	X				
Ponnusamy e Eswararaj (2023)	X				
Bianchi et al. (2003)	X				
Brown (1993)	X				
Dodero et al. (2019)	X				
Jomhari et al. (2024)	X				
Visaggio (2001)	X				
Medeiros et al. (2023)		X			
Baghdadi e Al-Bulushi (2015)			X		
Mateos et al. (2019)			X		
De Lucia et al. (2008)				X	
Noffsinger et al. (1998)				X	
Sneed (2000)				X	
Sneed (2009)				X	
Sneed e Verhoef (2019)				X	
Apaolaza e Vigo (2017)					X
Breitmayer et al. (2023)					X

4

Abordagem proposta para a Modernização de Sistemas Legados

Este capítulo apresenta a abordagem proposta para a modernização de sistemas legados, incluindo suas fases, a descrição do sistema de análise de logs desenvolvidos (seção 4.6) e cenários de sucesso e falha (seção 4.7).

A abordagem foi concebida para ser genérica e atender diversos tipos de sistema legado. Neste estudo, a abordagem foi aplicada a um site de estatísticas esportivas, definido na seção 5. Algumas análises e programas foram criados especificamente para o sistema legado. Essa deve ser uma prática comum em outras aplicações da abordagem, porém os conceitos principais serão utilizados em qualquer aplicação.

4.1

Introdução

A abordagem proposta para a modernização de sistemas legados é composta por quatro fases de desenvolvimento, conforme retratado na Figura 4.1.

Ao longo de todo o processo realizamos análises de log de interação com o sistema legado. Para isto desenvolvemos um sistema de análise de logs, descrito na seção 4.7.

Na primeira fase realizamos um primeiro diagnóstico do sistema legado, examinamos sua documentação, decidimos qual ferramenta de captura de *logs* será utilizada e consultamos o usuário principal para capturar seus principais



Figura 4.1: Abordagem de modernização proposta

problemas e objetivo. Em seguida, iniciamos a captura e análise dos *logs* (seção 4.2).

Na segunda fase, com os resultados iniciais da análise de *logs*, criamos e aplicamos os questionários. Em seguida caso necessário, criamos os roteiros de entrevistas com base nos resultados da análise de *logs* e resultado dos questionários (seção 4.3).

De posse de todos os resultados da segunda fase, continuamos com a análise de *logs* e combinamos os dados para a geração das recomendações de modernização no sistema e planejamos o desenvolvimento dos ajustes (seção 4.4).

Finalmente, implementamos os ajustes e incrementamos a análise de *logs* com a busca de *logs* relacionados aos ajustes. O desenvolvimento continua em paralelo ao processo de avaliação dos resultados, que podem gerar novos questionários e entrevistas e, por consequência, novas recomendações, retornando as fases anteriores, em um ciclo iterativo (seção 4.5).

4.2

Fase 1 - Coleta de dados

Como principal objetivo da primeira fase realizamos o diagnóstico inicial, conforme especificado na figura 4.2. A fase 1 começa com a análise da documentação do sistema, entrevista com o usuário chave e a análise do próprio sistema legado.

O diagnóstico inicial da abordagem consiste em entrevistar usuários-chave do sistema, analisar documentação existente, analisar documentação do código do sistema e obter uma visão inicial do sistema legado.

O usuário-chave do sistema pode ser a pessoa que tem maior conhecimento do funcionamento do sistema, a pessoa que seja proprietária do sistema ou o principal patrocinador do sistema.

Para a condução desta pesquisa, o usuário principal identificado foi o proprietário do sistema.

A identificação de documentação do sistema e análise do sistema legado são necessários para o funcionamento pleno da abordagem. Um sistema legado sem documentação inviabiliza a utilização da abordagem. Para ajudar nessa tarefa desenvolvemos um programa de extração de informações. Esse programa, específico para sistemas legados web, busca os dados específicos no HTML que serão listados nos logs de execução, extrai tags `<title>`, para gerar a documentação dos possíveis títulos de página HTML e dos elementos `<a href...>` e as URLs envolvidas contidas em cada página do sistema. O

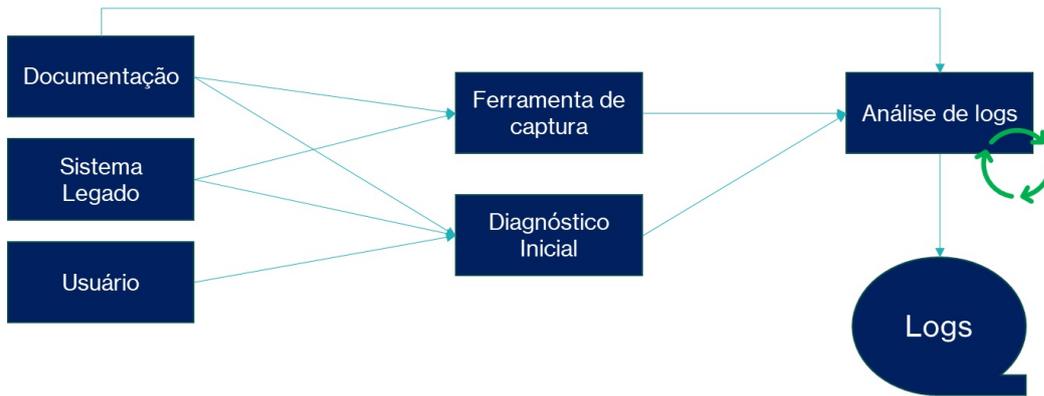


Figura 4.2: Abordagem de modernização - Fase 1: Coleta de dados

programa para extração de informação *Identificando_links.ipynb*, descrito no Apêndice B.1.8.

Uma documentação robusta é muito importante para a aplicação da abordagem, bem como a realização de uma entrevista com o usuário chave, para delinear os objetivos e o escopo inicial que irão embasar as análises iniciais de *logs*.

Com os objetivos e a documentação em mãos devemos decidir qual ferramenta de análise de *logs* utilizar. Em alguns casos, o próprio sistema legado pode ter alguma ferramenta já implementada. Caso contrário, devemos escolher uma ferramenta, semelhante ao Google Analytics¹.

Após a implantação e configuração da ferramenta de captura de *logs*, definimos a data da primeira coleta e a periodicidade da coleta. Para isso consideramos a quantidade de itens gerados e limitações da ferramenta de captura. Por exemplo, a versão gratuita do Google Analytics permite somente a extração de 5000 registros por download.

Por fim, iniciamos a coleta dos *logs* do sistema, a definição do que deve ser extraído vai variar dependendo do sistema e principalmente dos objetivos traçados para a modernização, uma sugestão para um sistema legado web seriam os logs relativos aos títulos acessados, as páginas acessadas, neste caso analisamos as URLs e o número de acessos total ao sistema, com isso realizamos as primeiras análises, buscando os primeiros indicativos de utilização do sistema legado, quais títulos/páginas são mais e menos acessados, número de usuários acessando o sistema.

A análise dos *logs* é dividida em duas fases: uma manual, de pré-processamento dos dados, e outra automática, fazendo uso dos programas que

¹<https://analytics.google.com>

desenvolvemos. Desenvolvemos inicialmente sete programas de análise de *logs*, são eles:

- Análise de títulos por dia - apêndice B.1.2
- Análise de títulos por hora - apêndice B.1.3
- Análise de acesso a páginas por dia - apêndice B.1.4
- Análise dos acessos ao sistema - apêndice B.1.5
- Análise de Títulos com pouco ou nenhum acesso - apêndice B.1.6
- Análise da tecnologia utilizada para acessar o sistema - apêndice B.1.7
- Gera referência - apêndice B.1.1

Este último programa gerado faz uma correlação entre as URLs acessadas e as consultas existentes no sistema, isso foi necessário porque uma mesma tela pode ser acessada por caminhos distintos e precisamos computar os acessos de forma distinta, na saída deste processo é gerado o arquivo *referencia.csv* que será citado em outros pontos dessa dissertação.

A abordagem flexibiliza o uso desses programas: somente o que for necessário deve ser executado. Não existe uma regra para a criação de novos programas de análise de *logs*, vai depender do caminho que as análises vão levar e da disponibilidade do log na ferramenta escolhida.

A extração dos *logs* na ferramenta é feita dentro de um período definido, portanto o pré-processamento dos dados consiste em concatenar os arquivos de *logs* extraídos da ferramenta com um arquivo de mesmo formato com todas as linhas, chamamos este de dataset principal, o formato de cada dataset vai depender dos campos de interesse que estão sendo capturados na extração.

Um ponto de atenção nestes datasets diz respeito a acentuação. Em determinadas instalações, os acentos podem se tornar um problema nos programas automáticos, seja na análise ou na exibição dos dados. Para isso foi criada uma macro no aplicativo Notepad++² que realiza ajustes de acentuação e transforma as vírgulas (,) em ponto-e-vírgula (;), para facilitar acesso a esses datasets através do aplicativo Microsoft® Excel.

Com os arquivos extraídos, realizamos o processamento dos dados através dos programas criados, conforme especificado no apêndice B. A execução dos programas é independente entre si. Não existe uma ordem para a execução, podendo até mesmo omitir alguma execução para dar foco em algum estudo mais urgente, de acordo com os objetivos da modernização do sistema legado.

Todos os programas têm uma configuração das datas de início e fim da análise, permitindo a delimitação de um período de tempo para a geração das

²<https://notepad-plus-plus.org/downloads/>

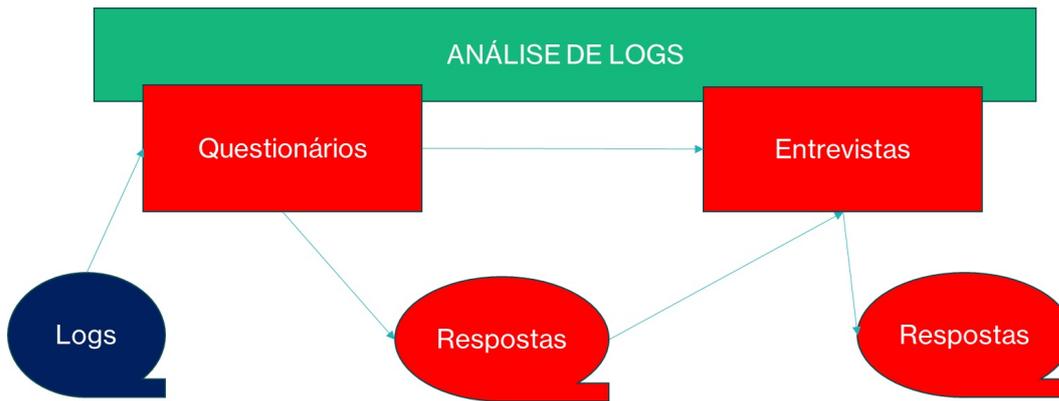


Figura 4.3: Abordagem de modernização – Fase 2

informações. Os gráficos e datasets de saída são salvos com essas datas como parte do nome dos arquivos, para permitir avaliar de forma incremental os dados coletados.

4.3

Fase 2 - Questionários e Entrevistas

A segunda fase da abordagem começa com a continuação da análise dos *logs* iniciada na primeira fase, conforme descrito na figura 4.3.

Conforme os resultados vão sendo coletados e interpretados, questões sobre a utilização do sistema legado vão surgindo e devem ser listadas para a confecção de questionários. Também podem servir de insumo para roteiros de entrevista.

Os tipos de questão que consideramos são:

- Porque uma área do sistema tem pouco ou nenhum acesso?
- Qual a relação entre as consultas do sistema com notícias relevantes atuais?
- Quando temos o maior número de acessos ao sistema?

Os questionários devem ser aplicados com os usuários principais do sistema. Podemos ter dois cenários de sistema legado. Quando temos um sistema legado aberto, como um site utilizado por diversos usuários, podemos incluir no próprio sistema um chamado para que os usuários respondam e ajudem a melhorar o sistema. Para os casos de sistemas legados mais controlados, dentro de uma rede interna de uma empresa, podemos enviar o questionário para usuários indicados e ter uma base de respostas mais bem estruturada devido ao maior conhecimento do usuário na utilização do sistema.

Os prazos para a coleta de respostas deve ser definido pelos analistas em comum acordo com o usuário principal do sistema, no início do processo e podem ser definidos por um período determinado ou pela quantidade de respostas para que tenhamos uma boa base de informações que nos ajude a complementar os dados coletados na análise dos *logs*, sugerimos que esses questionários sejam realizados em uma ferramenta online como google forms ou similar, para facilitar a coleta e compilação dos resultados.

Caso os questionários não resolvam todas as dúvidas ou se surgir algum fato diferente do que foi observado na análise de *logs*, geramos o roteiro de entrevistas com base nas análises iniciais dos *logs* complementadas com os dados do questionário. O que devemos buscar nessas entrevistas é como os usuários utilizam o sistema, quais as suas maiores dificuldades, quais os maiores problemas e frustrações enfrentadas e ideias que tenham para a melhoria no sistema.

4.4

Fase 3 - Recomendação de Ajustes

Com os resultados das análises de *logs*, dos questionários e das entrevistas, realizamos uma análise conjunta das informações, visando os objetivos e desafios listados no diagnóstico inicial, e preparamos as primeiras sugestões de modernização, preenchendo um documento de recomendações, este documento deve conter uma introdução contendo o histórico do sistema e sua arquitetura, uma breve descrição da coleta de dados e dinâmica de questionários e entrevistas, o público alvo do sistema, uma breve descrição da análise de logs, questionários aplicados e entrevistas realizadas e a lista de sugestões derivadas do processo. Na continuação devemos descrever o procedimento da análise de logs, com detalhes do tratamento dos logs, programas utilizados e o processamento das informações e uma lista com seus resultados, o mesmo deve ser feito para os questionários e entrevistas. Por fim as sugestões devem ser detalhadas, os resultados, após coletados, devem ser detalhados e a avaliação do usuário deve ser preenchida.

Em paralelo a esse processo, os *logs* continuam sendo coletados e analisados, conforme especificado na figura 4.4.

Após a criação do documento de recomendação planejamos o desenvolvimento das modificações sugeridas.

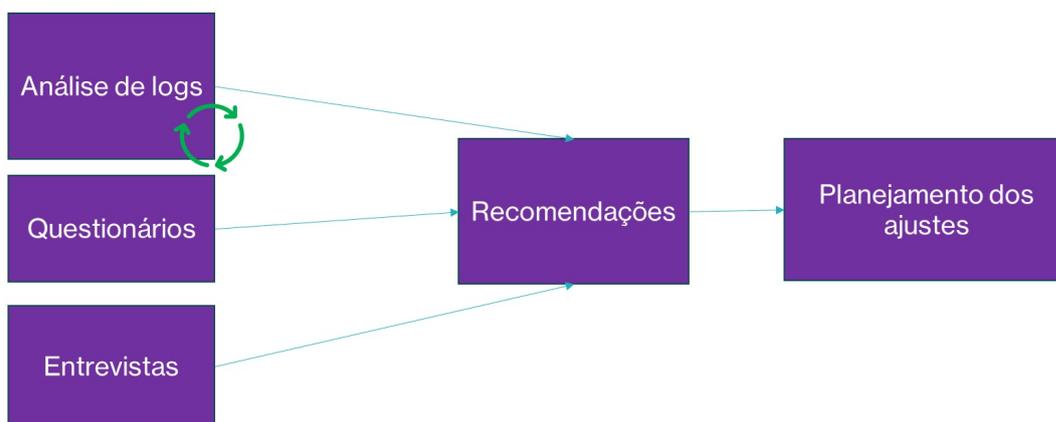


Figura 4.4: Abordagem de modernização – Fase 3 - Recomendação de Ajustes

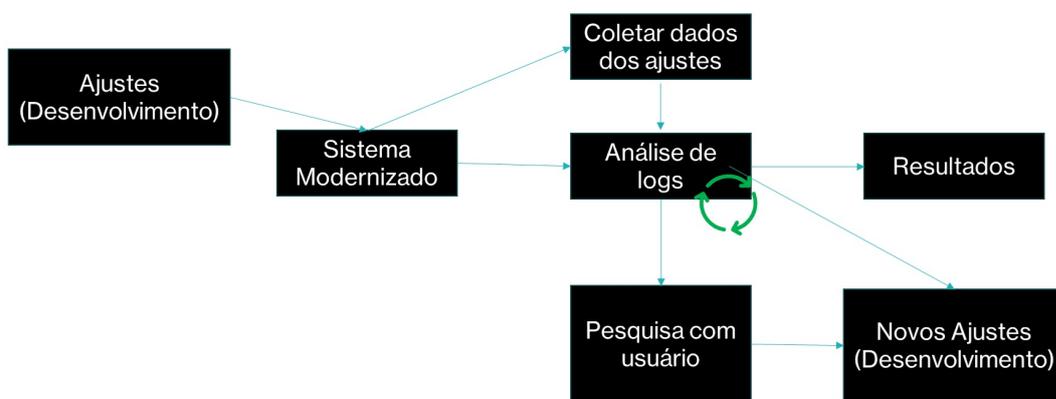


Figura 4.5: Abordagem de modernização – Fase 4 - Desenvolvimento e avaliação de ajustes

4.5

Fase 4 - Desenvolvimento e Avaliação

Nesta fase implementamos as recomendações, conforme especificado na figura 4.5.

A coleta de logs continua a ser realizada e novos relatórios começam a ser extraídos para que sejam verificados os ajustes que estão sendo implementados. Esses novos logs têm por finalidade apontar o que está acontecendo com a experiência do usuário dentro do sistema.

Nesta análise buscamos identificar os impactos das sugestões implementadas no sistema, os pontos a serem analisados vão variar dentro dos objetivos listados porém sugerimos analisar engajamento, número de acessos totais, número de acesso página modernizada como tópicos importantes para qualquer sistema.

O resultado da nova análise de *logs* pode gerar a criação de novos questionários e entrevistas para aumentar o entendimento sobre os ajustes realizados e avaliação dos resultados do processo de modernização do sistema.

Esse processo vai se retroalimentando e novos ajustes vão sendo identificados e depois implementados e avaliados.

4.6

Sistema de análise de logs

O sistema de análise de logs é composto por 2 fases, a primeira engloba a coleta e preparação dos dados e a segunda a execução dos programas para análise dos logs e geração de resultados, todo o processo deve ser guiado pelos objetivos listados para a modernização e pelo tipo de sistema legado que estamos trabalhando, para efeito dessa dissertação estamos usando um sistema legado web como base.

A primeira fase do sistema ainda está com processos manuais, processos que devem ser automatizados em uma próxima versão do sistema, e consiste em baixar os arquivos de *log* da ferramenta escolhida, na periodicidade especificada e concatenar com os arquivos principais, que serão a entrada do sistema, isso deve ser feito para cada arquivo de log escolhido para ser analisado, porém antes da concatenação com o arquivo principal o arquivo baixado deve ser verificado quanto a questão de acentuação, foi criado um processo semi-automático para resolver todos esses problemas descrito na seção 4.2, a concatenação dos arquivos deve ser feita manualmente utilizando um programa que consiga trabalhar com a extensão CSV, a estrutura de cada arquivo a ser baixado depende dos objetivos da modernização e será definida na ferramenta de extração de logs.

Após a concatenação dos arquivos entramos na fase de execução dos programas, todos os programas tem um filtro de data para delimitar o período de análise que desejamos executar, a ordem e a execução de cada programa depende dos objetivos de modernização, todos os gráficos gerados são salvos, o local de salva dos arquivos fica a cargo do desenvolvedor e cada gráfico tem as datas limites como diferenciador, desta forma todos os outputs gerados são salvos para futura referência e análise, estes outputs serão a principal fonte para a principal saída do sistema de análise de logs que são as sugestões de modernização, os programas criados são:

- Análise de títulos por dia - apêndice B.1.2 - Este programa busca nos *logs* de acesso as páginas do sistema web pela coluna do título da página e faz análises de quantidade de acessos por data e quantidade de acessos geral, faz filtros por palavras-chave, a serem determinadas no curso da análise dependendo do sistema legado e seus objetivos de modernização, identificando picos de acesso e trazendo informação que será confrontada com eventos externos acontecidos no mesmo período, além de trazer

- gráficos das títulos com mais acessos, títulos com menos acessos no geral e por data.
- Análise de títulos por hora - apêndice B.1.3 - Este programa busca nos *logs* de acesso as páginas do sistema web pela coluna do título da página e faz análises de quantidade de acessos por data e hora, faz filtros por palavras-chave, a serem determinadas no curso da análise dependendo do sistema legado e seus objetivos de modernização, identificando picos de acesso e trazendo informação que será confrontada com eventos externos acontecidos no mesmo período.
 - Análise de acesso a páginas por dia - apêndice B.1.4 - Este programa busca nos *logs* de acesso as páginas do sistema web pela coluna de url de acesso da página e faz análises de quantidade de acessos por data e quantidade de acessos geral, pode fazer filtros por palavras-chave, a serem determinadas no curso da análise dependendo do sistema legado e seus objetivos de modernização, identificando picos de acesso e trazendo informação que será confrontada com eventos externos acontecidos no mesmo período, além de trazer gráficos das urls com mais acessos, urls com menos acessos no geral e por data.
 - Análise dos acessos ao sistema - apêndice B.1.5 - Este programa verifica o número de acessos ao sistema por dia, engajamento total médio dos usuários por dia gerando gráficos das variações de acesso e engajamento que serão confrontados com eventos externos a procura de correlações.
 - Análise de Títulos com pouco ou nenhum acesso - apêndice B.1.6 - Este programa verifica os títulos que nunca tiveram acesso e os que tem muito pouco acesso, a quantidade de poucos acessos deve ser personalizada dentro da realidade de cada sistema legado, essa informação gerada deve ser utilizada como insumo para a criação dos questionários com a intenção de entender quais as razões do pouco ou nenhum acesso.
 - Análise da tecnologia utilizada para acessar o sistema - apêndice B.1.7 - Este programa identifica qual o dispositivo mais utilizado pelos usuários no acesso ao sistema, dependendo do sistema legado e seus objetivos podemos chegar a conclusões relacionadas ao design do sistema e responsividade, além de identificar de onde o usuário mais acessa o sistema, se foi via rede social, de forma direta, de um buscador, essa informação será útil para um melhor contato com o usuário na rede social, se isso fizer sentido para o sistema legado, criar mecanismos que facilite os buscadores encontrarem seu sistema.

- Análise Questionário - apêndice B.1.9 - Este programa compila os resultados informados do questionário, gerando gráficos que apoiam na tomada de decisão quanto a realização de entrevistas ou quanto a formulação de sugestões de modernização.
- Análise Atividades - apêndice B.1.11 - Este programa analisa os resultados de engajamento na URL, número de clicks e qualquer outra coluna que seja importante, essas colunas devem estar em linha com os objetivos de modernização do sistema legado e deve ser selecionada na ferramenta de coleta de dados, ele fornece gráficos que vão trazer informações importantes, um alto engajamento em determinada URL comparada com outras URLs pode indicar dificuldade de entendimento da página, processamento lento, um grande número de atividades como clicks dentro de um mesmo acesso a URL podem indicar problemas no design da tela que geram dificuldade de entendimento.
- Análise redes sociais - apêndice B.1.10 - Este programa compila o retorno gerado pelas postagens realizadas em redes sociais, nesta primeira versão somente a rede X foi contemplada devido a movimentação muito superior as demais redes sociais(Facebook e Instagram).

Lista de requisitos funcionais e não funcionais que devem ser verificados para a aplicação da abordagem. Esses requisitos foram definidos pela nossa experiência profissional e observações sobre processos de modernização que terminaram sem sucesso e com sucesso.

O sistema legado escolhido para a aplicação da abordagem atende a todos os requisitos através da sua documentação interna ao código e ao sistema de análise de *logs* escolhida para a coleta.

4.6.1

Requisitos Funcionais do Sistema de Análise de *logs*

- RF1 – O sistema deve ter mecanismos de buscas individuais ou em grupo a fim de reunir títulos de página do mesmo assunto e assim permitir uma melhor validação dos resultados.
- RF2 – A ferramenta de coleta de *logs* deve ser capaz de gerar relatórios no formato CSV.
- RF3 – O sistema deve ser capaz de gerar a análise de *logs* para qualquer período desejado.
- RF4 – O sistema deve guardar todos os gráficos gerados com a distinção do período consultado para posterior análise de resultados e comparações.

RF5 – O sistema deve conseguir varrer os HTMLs existentes para identificação de links via a tag “a” da linguagem.

RF6 – O sistema de conseguir identificar os links existentes codificados através do comando “url for”.

RF7 – O sistema deve ser capaz de identificar e listar todas as páginas com determinado número de acessos únicos.

4.6.2

Requisitos Não Funcionais do Sistema de Análise de logs

RNF1 Escalabilidade – O sistema deve permitir a criação de análises de novos parâmetros que sejam solicitados.

RNF2 Confiabilidade – O sistema deve ser confiável e atender aos requisitos.

RNF3 Usabilidade – O sistema deve ser de fácil entendimento e uso.

4.7

Cenário de Aplicação da Abordagem Proposta

4.7.1

Cenário de Sucesso 1

Primeiro cenário do sistema legado que proporciona condições de aplicação da abordagem com sucesso.

4.7.1.1

Situação do Sistema Legado:

1. O sistema legado possui documentação robusta com títulos únicos criados em todas as páginas e com significado claro.
2. As páginas sendo acessadas possuem urls distintas, as funções ou classes são separadas por grupo relacionado, e tudo isso é documentado de forma robusta e completa.
3. Um sistema de coleta de dados de uso como o Google Analytics ou análogo está instalado e gera todos os relatórios apresentados com sucesso.

4.7.1.2

Desenvolvimento da Análise

A análise transcorre sem problemas, respeitando as frequências de recuperação e análise de *logs*, tempo de exposição do questionário e entrevistas com usuários-chave realizadas com sucesso.

4.7.1.3

Resultados

As recomendações de redesign são elaboradas, implementadas e avaliadas com sucesso. O sistema legado é modernizado e transformado em uma interface inteligente sem perda de nenhuma regra de negócio, trazendo conforto e melhorando a experiência do usuário no uso do sistema. Como resultado, o reprojeto aumenta a produtividade, engajamento, exposição da marca e, quando pertinente, a geração de receita.

4.7.2

Cenário de Sucesso 2

Segundo cenário do sistema legado que proporciona condições de aplicação da abordagem com sucesso.

4.7.2.1

Situação do Sistema Legado:

1. O sistema legado possui documentação robusta com títulos criados em todas as páginas e com significado claro. Porém, alguns títulos estão duplicados e precisam ser distinguidos de uma outra forma.
2. As páginas são acessadas com códigos distintos e as funções ou classes são separadas por grupo relacionado. Porém, não há uma documentação robusta, e a criação do arquivo referencia.csv não é possível diante do cenário encontrado.
3. Com a execução o programa de identificação de links, é possível fazer a geração do arquivo referencia.csv (seção 4.2) com uma boa acurácia e utilizá-lo para a análise de acesso às páginas do sistema.
4. Um sistema de coleta de dados de uso como o Google Analytics ou análogo está instalado e gera todos os relatórios apresentados no Manual com sucesso.

Desenvolvimento da Análise

Precisa ser gasto um tempo maior para a criação do arquivo de referencia.csv (seção 4.2).

De tempos em tempos, é necessária uma revisão para a identificação de novas páginas adicionadas, identificação dos títulos duplicados e criação de lógica dentro dos notebooks para tratar o problema de identificação das consultas, afastando qualquer ruído na análise dos *logs*, respeitando as frequências de recuperação e análise de *logs*. Tempo de exposição do questionário e entrevistas com usuários-chave realizadas com sucesso.

4.7.2.2 Resultados

Com mínima interferência no resultado final, fazemos as recomendações. Implementamos e avaliamos com sucesso.

Modernizamos o sistema legado e realizamos essa transformação sem perda de nenhuma regra de negócio.

Trazendo conforto e melhorando a experiência do usuário no uso do sistema, aumentando produtividade, engajamento, exposição da marca e geração de receita.

4.7.3 Cenário de Falha 1

Primeiro cenário do sistema legado que não proporciona condições de aplicação da abordagem com sucesso.

4.7.3.1 Situação do Sistema Legado:

1. O sistema legado não possui documentação robusta e algumas páginas não possuem título. Algumas páginas possuem títulos genéricos, que não conseguem produzir uma distinção sobre o assunto que está sendo abordado na página ou consulta.
2. As páginas sendo acessadas possuem códigos duplicados e as funções ou classes estão misturadas, tornando impraticável dividir os acessos em grupos. A criação do arquivo referencia.csv não é possível diante deste cenário.
3. Indefinição de grupos e subgrupos de consultas e páginas
4. Mesmo com a execução do programa de identificação de links, ainda não é possível a criação do arquivo referencia.csv.

5. Um sistema de coleta de dados de uso como o Google Analytics ou análogo está instalado e gerando alguns dos relatórios apresentados no Manual com sucesso, devido à falha de definição dos títulos.

4.7.3.2

Desenvolvimento da Análise

A definição clara dos títulos e a geração de um arquivo de referência, mesmo que incompleto, são as duas condições para que a abordagem traga algum ganho e gere alguma sugestão. Neste cenário, a análise não é realizada.

4.7.3.3

Resultados

A análise não é realizada, nenhuma sugestão é possível de ser feita. Em um caso como este, recomendamos partir para uma modernização tradicional, na qual alguma ferramenta extraia regras de negócio do sistema legado e o design seja feito completamente do início. Nesses casos, em geral não vale a pena investir em ajustar o sistema legado nem criar uma documentação.

4.7.4

Cenário de falha 2

Segundo cenário do sistema legado que não proporciona condições de aplicação da abordagem com sucesso.

4.7.4.1

Situação do Sistema Legado

1. O sistema legado possui muitos títulos em duplicidade ou genéricos tornando a identificação da navegação muito confusa porém possível.
2. As páginas sendo acessadas possuem códigos duplicados, e as funções e classes estão misturadas. A criação do arquivo referencia.csv não é possível diante do cenário encontrado.
3. Existe alguma definição entre grupos e subgrupos.
4. Com a execução do programa de identificação de links, é possível fazer a geração do arquivo referencia.csv, mas com uma péssima acurácia, o que o torna muito ineficiente na análise de páginas acessadas.
5. Um sistema de coleta de dados de uso como o Google Analytics ou análogo está instalado e gerando alguns dos relatórios apresentados no Manual com sucesso, devido à falha de definição dos títulos.

4.7.4.2

Desenvolvimento da Análise

Alguma análise dos *logs* é feita, porém os resultados são incapazes de trazer qualquer tipo de informação que possa ser utilizada para que seja feita uma sugestão de modernização com segurança.

4.7.4.3

Resultados

A aplicação da abordagem não traz os resultados esperados e o projeto de modernização não pode ser concluído. Uma solução seria fazer um projeto de modernização tradicional, no qual alguma ferramenta extraia regras de negócio do sistema legado e o design seja feito completamente do início. Nesses casos, em geral não vale a pena investir em ajustar o sistema legado nem criar uma documentação.

4.8

Considerações

Este capítulo descreveu a abordagem proposta em detalhes, para que seja possível que outras pessoas a utilizem.

O próximo capítulo apresenta um estudo da aplicação da abordagem.

5

Estudo de aplicação da abordagem proposta

Este capítulo apresenta o estudo preliminar que desenvolvemos para a análise dos dados de navegação do site Fluzao.xyz com base nos *logs* do servidor e em um questionário aberto definido na Seção A.2, disponível através do próprio site por 30 dias de 01/10/2023 até 31/10/2023, conforme especificado nas seções 4.2 e 4.3.

5.1

Objeto de estudo

Para realizar a prova de conceito da abordagem, escolhemos o site `fluzao.xyz`. Trata-se de um sistema próprio que foi iniciado no final dos anos 90, em uma aplicação COBOL com arquivos VSAM.

Conforme a figura 5.1, a primeira versão do sistema era um sistema local que tinha bastante informação sobre os atributos principais de um jogo de futebol, adversário, árbitro, estádio, campeonato, jogador, técnico e resultado, mas que mantinha as informações restritas.



Figura 5.1: Histórico do site `fluzao.xyz`

No ano de 2007 com a consolidação da internet surgiu a ideia de disponibilizar essas informações para o público em geral e portanto a primeira versão do site em PHP foi colocada no ar já no ano de 2008. Com a transformação dos arquivos em banco de dados, muitas informações começaram a ser cruzadas, abastecendo cada vez mais o público com dados interessantes. Para garantir a robustez destes dados, buscamos completar os dados faltantes e, com a participação de funcionários do clube tiramos fotos, das súmulas oficiais e completamos a base.

A partir daí, novas consultas foram surgindo e, mesmo sem o devido crédito, muitos números falados em transmissões esportivas do Fluminense eram e ainda são extraídos deste sistema.

Em 2020, tentando adequar o sistema com novas tecnologias, houve uma grande reformulação trocando o PHP/HTML por uma arquitetura em Python-Flask/HTML/Bootstrap 4, seguida por uma nova grande reformulação dos dados, com a publicação de um livro chamado Almanaque Tricolor, que passou a ser considerado pelo clube como a principal fonte de dados.

O sistema já contava com mais de 6000 jogos cadastrados, 69 tabelas mysql com 22,1Mb de dados, 580 consultas diferentes aos dados e disponíveis para o usuário. Porém tudo isso era mostrado em um ambiente sem interação eficiente com o usuário, onde o próprio usuário era o responsável por buscar a informação que ele queria ver. Surgiu então a terceira grande reformulação, utilizando a abordagem de modernização de sistemas legados para sugerir mudanças que melhorassem a experiência do usuário e trouxessem um sistema mais vibrante e moderno, sem que as regras de negócio das consultas fossem perdidas.

5.2

Planejamento do Estudo Preliminar

Conforme especificado na seção 4.2 da abordagem, realizamos um levantamento dos desafios de modernização junto aos usuários principais. Além disso, observamos a existência de algumas características antes de iniciar o processo de análise, conforme os possíveis cenários de aplicação da metodologia descritos na Seção 4.7. Essas características são baseadas em experiência própria e observações feitas em outros processos de modernização de sistemas legados sobre o que seria mais adequado para realizar a análise e são a base da abordagem proposta:

1. Verificamos que o sistema estava bem documentado. Como estamos trabalhando com um site então é desejável que as páginas do sistema tivessem um título explicativo, pois isso será a principal fonte de referência para a compreensão do uso do sistema. Com títulos bem formatados e sincronizados com as informações apresentadas, pudemos analisar os caminhos percorridos pelo usuário, suas preferências e o que não estava sendo acessado.
2. Verificamos que havia um mapeamento dos caminhos do sistema, para que pudessemos diferenciar funcionalidades/consultas que terminavam na mesma *landing page* e que possuíam o mesmo título. Caso isso não existisse poderíamos utilizar o programa de extração de urls listado na Subseção B.1.8, programa *identificando_links.ipnyb*.

3. Instalamos o analisador Google Analytics¹, pois o sistema não possui gerador de *logs*.

Vale observar que o sistema não requer que os usuários se identifiquem, o que impediu a extração de informação de navegação e preferências de grupos distintos de usuários.

5.2.1

Objetivo

O objetivo do estudo foi identificar os principais padrões de navegação dos usuários do site através da análise de *logs* e do questionário que foi disponibilizado na página principal do site, alinhando as análises, descobertas e sugestões com os objetivos e desafios do diagnóstico inicial.

Durante a primeira análise identificamos alguns objetivos secundários, tais como listar as consultas com pouco ou nenhum acesso, identificar qual o principal dispositivo de acesso ao sistema e verificar como se comportavam os acessos durante a realização de uma partida.

5.2.2

Procedimento

De acordo com a abordagem, dividimos o procedimento em Diagnóstico Inicial; Coleta e análise dos dados de *logs*, questionários e entrevistas; Elaboração das recomendações de modernização.

5.2.2.1

Diagnóstico Inicial

Iniciamos a aplicação da abordagem fazendo um diagnóstico inicial que consistiu em uma inspeção no site www.fluzao.xyz, buscando informações das tags html, além de realizarmos a verificação da documentação existente, a identificação das *urls* e das expectativas do usuário-chave, chegando aos seguintes desafios, conforme figura 5.2:

1. Tornar o sistemas mais interativo
2. Integrar as diversas áreas do sistema
3. Engajar mais o usuário
4. Construir canal de comunicação entre o sistema e seus usuários
5. Monetizar sem prejudicar a experiência do usuário

¹<https://analytics.google.com>



Figura 5.2: Desafios de Modernização

5.2.2.2

Coleta e análise dos dados de *logs*, questionários e entrevistas

A análise de *logs* tem como objetivo principal analisar os caminhos percorridos pelo usuário no sistema, identificando as funções mais procuradas, as menos procuradas, as nunca procuradas, quais as preferências dos usuários no sistema, qual dispositivo (smartphone, desktop, tablets) é o mais utilizado para acessar o sistema, como o usuário inicia o acesso ao sistema (site de busca, redes sociais, acesso direto), e de quais cidades/estados vem a maior parte de seus acessos, analisando os picos de acesso a certas informações.

No nosso estudo, buscamos as informações dos títulos das páginas por data e por data/hora, das páginas por data, acesso de usuários, acesso de usuários por cidade e por estado, engajamento, qual o dispositivo utilizado e como iniciou o acesso, porém existe a flexibilidade de buscar novas informações ou até mesmo excluir alguns desses datasets de acordo com as necessidades.

Dependendo da natureza do sistema legado, a análise de *logs* deve fornecer informações de navegação divididas por perfil de usuário, identificar as preferências do usuário através de sua navegação e isolar os principais usuários de diferentes perfis para encontrar padrões e interesses de navegação. No nosso estudo, isso não foi realizado por não termos os usuários realizando login no sistema, mas entendemos ser um ponto importante a ser considerado no futuro.

A frequência da coleta dos dados vai depender da utilização do sistema. O Google Analytics tem uma limitação de 5000 linhas de *log* por relatório; dessa forma, devemos avaliar a melhor frequência de coleta com base na frequência de acessos ao sistema. Para o caso de estudo, determinamos uma frequência semanal para todos os relatórios, com a coleta sendo realizada toda segunda-feira, desde o dia 03/08/2023.

A análise inicial dos registros serviu de guia na criação dos questionários e entrevistas que foram aplicados na fase inicial da pesquisa.

Foram elaborados 13 programas de análise dos dados de *logs* e questionário provendo datasets e gráficos de saída que auxiliam na criação das sugestões

de modernização. O material elaborado para o estudo se encontra no Apêndice A.

O período de tempo em que o questionário fica disponível pode variar e depende do tempo que temos para realizar a pesquisa e a modernização do sistema, no nosso caso, determinamos um período de um mês para a resposta dos usuários.

Neste estudo preliminar, não conduzimos entrevistas, as informações dos *logs* e questionário se mostraram suficientes para a elaboração das primeiras sugestões de modernização.

Todo esse processo compreende a primeira e segunda fase da abordagem e os seus outputs foram a base para a elaboração das sugestões de modernização na terceira fase da abordagem.

5.2.2.3

Elaboração das recomendações de modernização

Depois de executadas as fases anteriores, coleta de dados, análise de *logs* e questionários, compilamos os dados e projetamos ajustes e alterações ao design atual, sistemas de recomendação e outros. Essas ideias foram documentadas em um formulário e analisadas pela equipe que buscava uma solução adequada.

As recomendações que foram idealizadas tendo como base a análise dos dados foram:

1. Criação de um sistema de recomendação de consultas por conteúdo substituindo a tela inicial estática antiga do sistema - tendo como base os *logs* de navegação que indicam a preferência dos usuários pelas notícias mais recentes, identificamos que o maior número de acessos está sempre concentrado no próximo adversário, nos jogadores atuais, em notícias atuais que impactam o clube, conforme demonstrado na seção 5.3.1. Essa recomendação se relaciona diretamente com os objetivos de maior engajamento do usuário, tornar o sistema mais interativo e integração das áreas do sistema.
2. Canal de comunicação nas redes sociais e identificação das novas consultas na tela principal - Utilizando o questionário e a identificação de onde se originam os acessos ao sistema ficou evidenciado que um canal de comunicação as redes sociais seria importante na expectativa de alavancar o número de acessos. Já a velocidade de crescimento de novas consultas implementadas pela análise de acessos às páginas do sistema indicou que uma comunicação sobre novidades criadas no sistema podia ajudar

na divulgação e consumo de novos dados relevantes para o momento do clube.

3. Mudança nas telas de consulta dos atributos principais do jogo (adversários, árbitros, estádios, campeonatos, jogadores e técnicos) - A análise de *logs* indicou que usuários nem sempre finalizam as suas consultas. Conforme descrito na seção 5.3.2, a divisão dos atributos entre recentes e antigos visa facilitar a busca, visto que, segundo a análise dos *logs*, o principal interesse dos usuários é sobre os dados mais recentes.
4. Identidade visual do site e identificação dos links - O questionário e análise de *logs* indicaram que a seção de gráficos do sistema tinha muito pouco acesso. No questionário 60% dos usuários disseram nunca acessar os gráficos, essa informação foi corroborada pela análise de *logs*, que indicou que 50% das páginas com menos acessos eram páginas de gráficos. 30% disseram não usar os links por não saberem onde eles estavam. Essa informação vai de encontro à análise de *logs*, que indica que as 9 páginas com mais acessos são páginas de links; porém a quantidade de usuários identificada no questionário indica que uma melhor visualização poderia proporcionar uma melhor experiência para todos os usuários.

5.3

Resultados do Estudo Preliminar

5.3.1

Análise de *logs*

Os *logs* são coletados toda segunda-feira, englobando o período da segunda anterior até o domingo. Em seguida, esses relatórios são incorporados nos datasets principais de cada análise. Durante a semana, os *scripts* listados no Apêndice B são executados e os resultados continuamente analisados para a geração das sugestões de modernização.

Por meio da análise de *logs*, identificamos padrões de navegação para nortear a definição de ajustes no design do sistema, tais como quais telas deveriam ser alteradas para proporcionar uma melhor experiência ao usuário, aumentando o tempo de navegação e engajamento e aumentando o número de acessos; quais telas não eram muito acessadas e quais links diretos eram utilizados.

Com os dados iniciais, estabelecemos o objetivo principal do questionário: identificar as preferências e dificuldades dos usuários do sistema, como eles entendem determinadas funções do sistema, por que não acessam certas

áreas do sistema que trazem mais valor que outras, se eles estão abertos a possibilidades específicas de mudança que podem ser propostas, o que lhes agrada na forma como o sistema funciona, o que lhes falta e o que o sistema pode proporcionar.

O principal ponto identificado na utilização do sistema de 03/08/2023 até 02/08/2024 foi a preferência destacada dos usuários pelos temas que estão ocorrendo no momento. Estamos analisando um site de estatísticas esportivas² e há uma correlação muito forte dos acessos com as notícias mais recentes e relevantes para o clube. Seguem abaixo alguns exemplos que ilustram este comportamento de acessos.

1. Acessos à página da Sociedade Esportiva Palmeiras – Um jogo aconteceu no dia 05/08/2023, e podemos perceber que a maior parte dos acessos ocorreu entre 03/08/2023 e 06/08/2023. Outro jogo aconteceu no dia 03/12/2023 e, novamente, a maioria dos acessos ocorreu nos dias entre 01/12/2023 e 03/12/2023, conforme visto na Figura 5.3.
2. Acessos à página do Sport Club Internacional – dois jogos aconteceram em 27/09/2023 e 04/10/2023 e, no período próximo ao jogo (3 dias antes e 1 dia depois), este adversário foi o mais acessado. Além da proximidade dos jogos, no dia 31/08/2023 o Sport Club Internacional se tornou o próximo adversário da Libertadores 2023. Foi então que atingimos um pico de acessos às páginas relacionadas a um clube que não fosse o próximo adversário, tornando clara a preferência do usuário por informações mais atuais, conforme visto também na Figura 5.3.
3. Importância dos adversários tradicionais (três outros grandes times da cidade do Rio de Janeiro): Quando chegamos perto do jogo com um deles, temos um pico de acessos, ainda maior que os demais adversários (de outras localizações). Além disso, mesmo sem jogos a serem disputados em um futuro próximo sempre temos um número considerável de acessos, conforme visto na Figura 5.4.
4. Quando o árbitro Wilmar Roldan foi designado a disputar a final da Libertadores 2023, a página destinada a ele atingiu o pico de acessos. Esse pico continua sendo até hoje o recorde de acessos a uma página desde o dia 03/08/2023, conforme visto na Figura 5.5.
5. O ex-goleiro do Fluminense Paulo Goulart faleceu no dia 15/09/2023. Nesse dia, a página destinada a ele teve um pico de acesso, conforme visto na Figura 5.6.

²<https://fluzao.xyz>

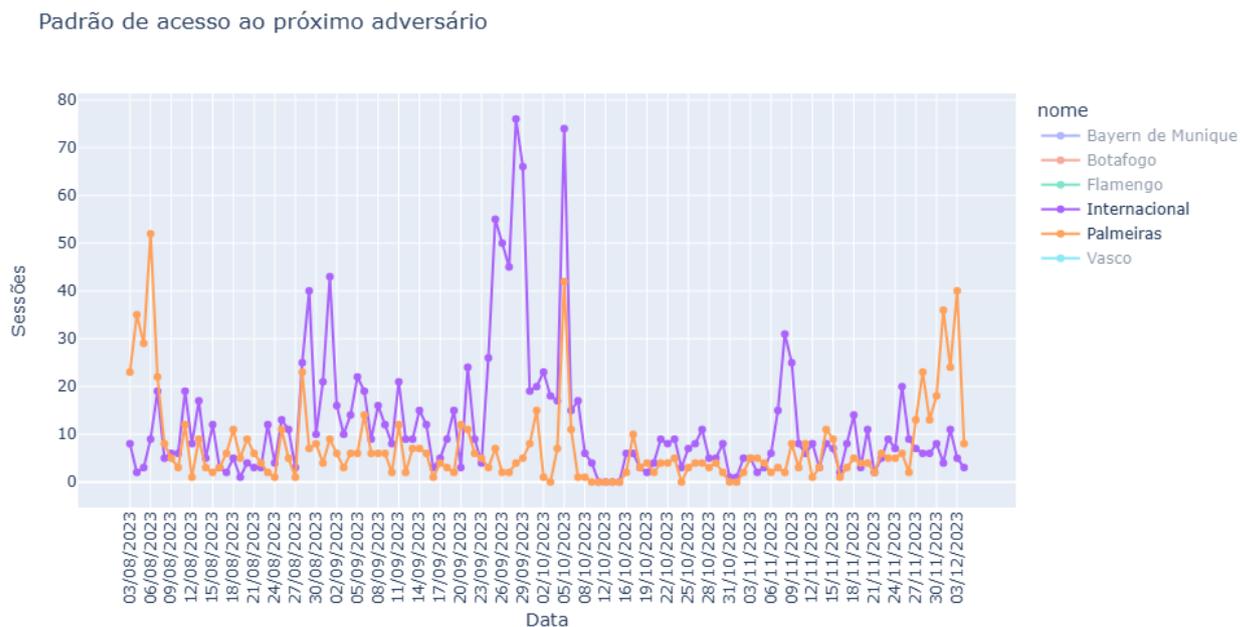


Figura 5.3: Acessos às páginas de Inter e Palmeiras

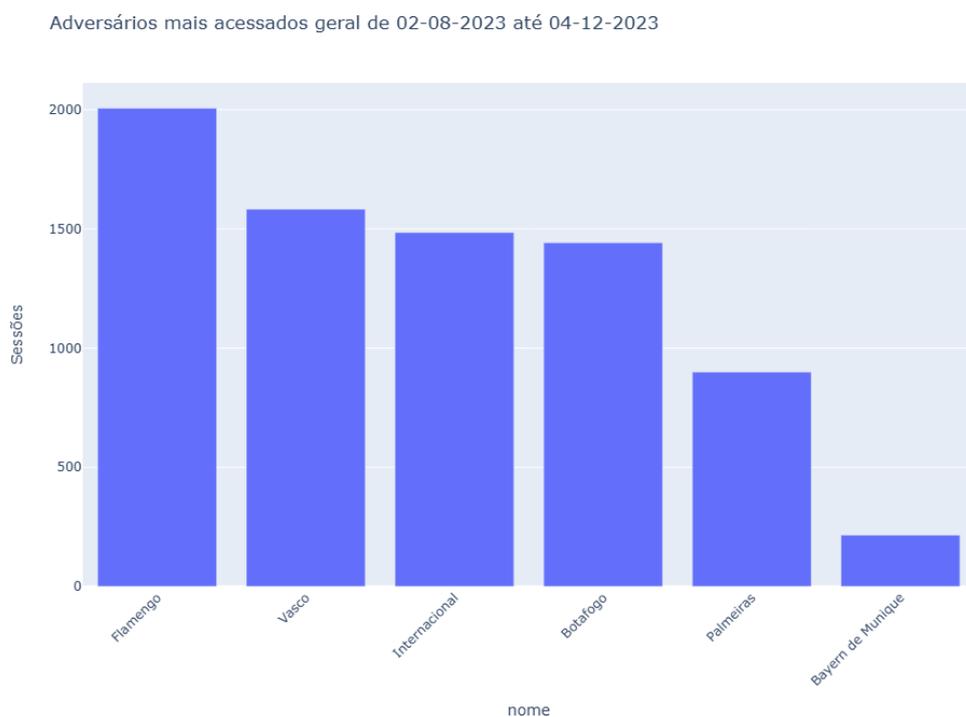


Figura 5.4: Acessos totais aos principais adversários

- Um exemplo diferente que demonstra que eventos fora do clube também podem impactar o interesse pelas informações foi a morte do jogador alemão Beckenbauer em 08/01/2024. Nesse dia, tivemos um dos maiores picos de acesso a informações relacionadas ao Bayern de Munique, clube

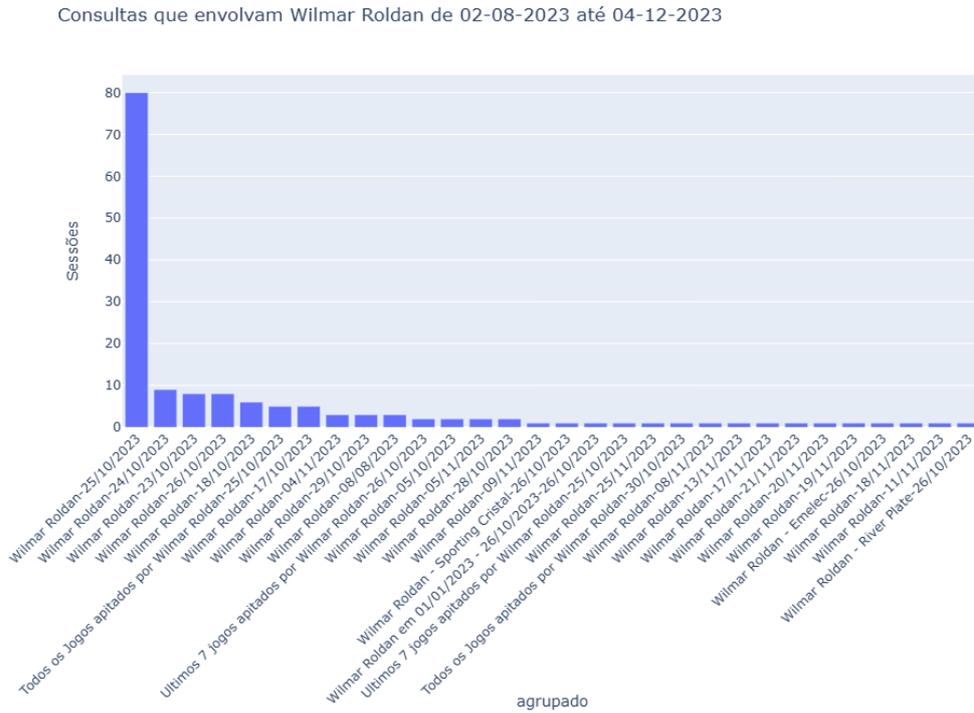


Figura 5.5: Acessos totais ao árbitro Wilmar Roldan

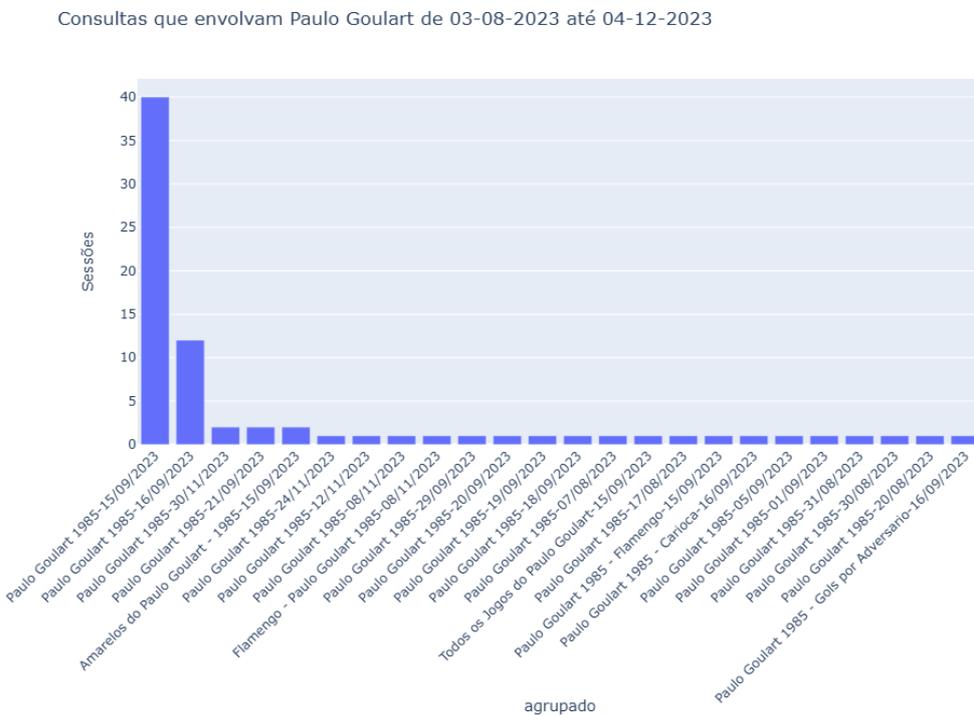


Figura 5.6: Acessos totais ao goleiro Paulo Goulart

pele qual Beckenbauer jogou sua única partida contra o Fluminense, em 1975, conforme visto na figura Figura 5.7

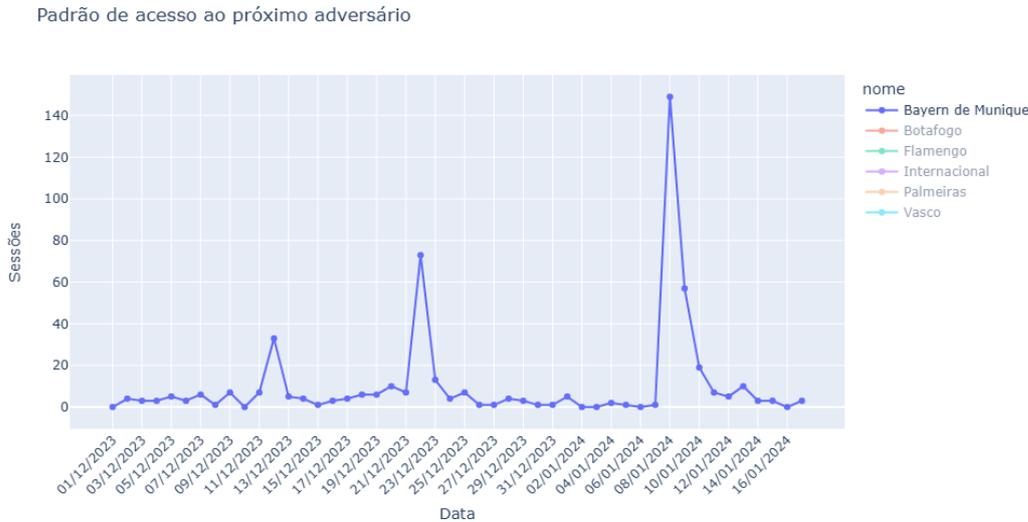


Figura 5.7: Acessos totais ao Bayern Munique

5.3.2 Questionário

Responderam ao questionário 42 pessoas de diversos perfis: jornalistas, torcedores, historiadores de futebol.

O questionário foi colocado à disposição dos usuários na página principal do site por 30 dias iniciando em 01/10/2023 até 31/10/2023 e foi respondido por 42 usuários. As respostas coletadas estão disponíveis no Apêndice A e foram são combinados com os resultados da análise de *logs* para a geração das sugestões de modernização presentes a Seção 5.4.

No questionário, listado na Seção A.2 criado para o site Fluzao.xyz, 80% dos usuários entendeu que seria muito interessante um sistema de recomendação, conforme Figura 5.9, trazendo as informações mais atuais sobre o clube. O sistema de recomendação por conteúdo foi o escolhido porque a recomendação por meio de filtragem colaborativa implicaria a inclusão de um perfil e criação de logins, o que foi recebido com menos entusiasmo, 70% dos usuários não gostaram da ideia da criação de um login, provavelmente pelo fato de compartilhar informações e pela possibilidade de ter algum custo financeiro futuro.

Identificamos também que os acessos são feitos 50% por telefones celulares e 50% por desktops, conforme Figura 5.8, e que os acessos são originados em buscadores ou por acesso direto. Os mecanismos de busca das consultas não representam nenhum tipo de desafio e agradam a maior parte dos usuários.

Um recurso importante são os links diretos entre páginas. Por exemplo, quando jogos são listados no site, são criados links para o adversário, árbitro,

Em qual plataforma você mais acessa o site?

38 respostas



Figura 5.8: Tecnologia usada para acessar site

Qual a sua avaliação sobre receber sugestões de consultas relevantes sobre assuntos atuais envolvendo o Fluminense?

38 respostas

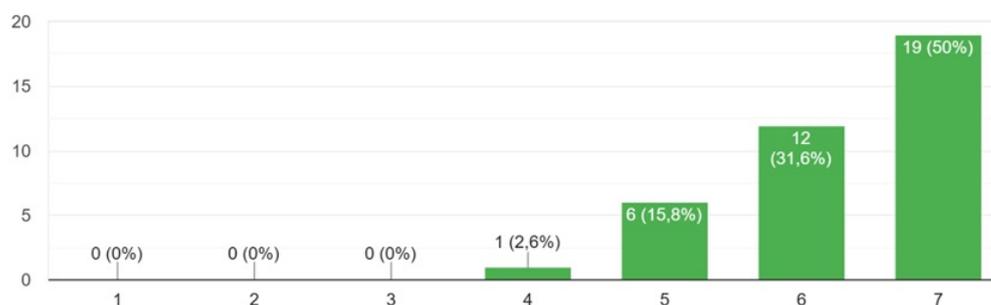


Figura 5.9: Avaliação sobre implementar sistema de recomendação

estádio, campeonato e o detalhe do jogo. No entanto, a maioria dos usuários ainda não sabe que tal ferramenta existe, o que indica que o site deve sinalizar melhor os seus atalhos.

Finalmente, a seção de gráficos não é acessada por mais da metade das pessoas que responderam o questionário. Em contrapartida, quem acessa as informações dos gráficos não relatou dificuldade na navegação.

Através das respostas ao questionário, os principais resultados foram:

- Confirmação da preferência dos usuários pelas informações mais recentes envolvendo o Fluminense, corroborando o que a análise de *logs* já tinha indicado.
- Confirmação que a seção de gráficos não era acessada conforme aponta-

mento da análise de *logs*.

- Pelo questionário, os links disponíveis passam despercebidos por 40% dos respondentes; no entanto, na análise de *logs*, identificamos que os cliques nesses links representam um percentual significativo de 80% das páginas acessadas.

5.4

Sugestão de Modernização

Esta seção lista as sugestões de modernização, estabelecendo as conexões com o que foi encontrado e como isso ajuda a responder as questões de pesquisa, conforme especificado nas seções 4.4 e 4.5.

Após a compilação de todo o material de análise de *logs* e os dados de resposta dos questionários, elaboramos as sugestões de modernização do sistema legado para melhorar a experiência do usuário em linha com os objetivos identificados no diagnóstico inicial, descrito na seção 4.2.

Geramos um documento listando as recomendações geradas e detalhes para o desenvolvimento e motivação (seção A.4). Vamos revisitar esse documento sempre que tivermos novos *insights* e avaliações das sugestões de recomendação.

5.4.1

Sistema de recomendação por conteúdo

A criação de um sistema de recomendação por conteúdo substituindo a página principal do sistema foi baseada em duas informações coletadas. No diagnóstico inicial do sistema, identificamos que o sistema não tem perfil de usuário, ou seja, não é possível seguir os passos individuais dos usuários. Mais de 60% das respostas do questionário sobre a criação de um perfil foram negativas, o que não recomenda a criação de um sistema de login nesse primeiro momento, afastando a possibilidade de criação de um sistema de recomendação por filtragem colaborativa. Assim, sugerimos a criação um sistema de recomendação por conteúdo com base nas informações mais atuais envolvendo o clube.

A característica de uso mais importante identificada pela análise dos *logs* é o interesse primário dos usuários pelas informações mais atuais envolvendo o clube. Isso foi confirmado com respostas unânimes sobre o mesmo interesse no questionário. Atendendo aos objetivos de maior engajamento do usuário, decidimos tornar o sistema mais interativo e construir um canal de comunicação com o usuário, relacionando isso com o grande interesse em informações atuais.

Com a revelação dos primeiros resultados indicando a preferência dos usuários no consumo das notícias relacionadas aos assuntos mais atuais, o analista de *logs* responsável pelo estudo sugeriu a criação de um sistema de recomendação com foco no próximo jogo, e um segundo sistema de recomendação com base em outras notícias que impactam o dia a dia da torcida.

Em 27/01/2024, o sistema de recomendação foi implantado como página principal do sistema. O funcionamento do sistema começa com o cadastro da notícia mais impactante do momento; um exemplo é o cadastro do próximo jogo. Nesse caso, é feito um cadastro contendo os dados do próximo adversário como uma das recomendações e mais cinco consultas relacionadas ao assunto principal, conforme a Figura 5.15. Durante o mês de março de 2024, implementamos o sistema de recomendação secundário, Figura 5.16, visando a atingir outros assuntos importantes com um *link* para consultas relacionadas ao assunto listado como secundário trazendo atalhos para essas consultas.

O formato de 6 consultas na tela principal com a consulta principal fixada na primeira posição e as outras 5 consultas relacionadas se revezando na tela foi definido pelo pesquisador e analista-programador e foi implementado com a opção de avaliação das consultas na tela. Além disso, foram incluídas no segundo questionário perguntas sobre a preferência do usuário entre as 2 telas iniciais e sobre a avaliação da nova tela inicial e sistema de recomendação. Quando necessário, foram feitas entrevistas com os usuários que preferiram ou deram avaliações negativas das nova tela.

5.4.2

Criação de canais de comunicação via rede social

Com o objetivo de atender o objetivo de uma melhor comunicação com o usuário, criamos contas próprias nas principais redes sociais da atualidade (X³, Instagram⁴ e Facebook⁵). Essa sugestão teve como base as respostas positivas de 80% dos usuários recebidas no questionário e na identificação através dos *logs* que os acessos ao sistema oriundos de redes sociais não chegavam a 2%.

5.4.3

Mudança na forma de consulta dos atributos primários

Os atributos primários do sistema legado são os adversários, estádios, árbitros, campeonatos, jogadores e técnicos envolvidos em uma partida de futebol. Eles são tratados dessa forma no sistema por terem sido os primeiros dados armazenados e explorados pelo sistema, ainda na década de 90.

³<https://twitter.com/EstatisticasFlu>

⁴<https://www.instagram.com/estatisticasfluminense>

⁵<https://www.facebook.com/estatisticas.fluminense>

Alinhando os objetivos de maior engajamento do usuário e tornando o sistema mais interativo, as pesquisas dos atributos primários eram dificultadas pelas quantidades de informações envolvidas. Um exemplo são os 1249 árbitros: a análise de *logs* revelou que 20% das sessões que chegavam na consulta de árbitro não completavam a ação.

Com esses dados de abandono, quantidade grande de informações juntas em uma única combo e mais a informação de que os usuários preferem consultar o que é mais atual, fizemos a sugestão de particionar os atributos primários em Recentes e Antigos, gerando, por exemplo, uma diminuição da lista de árbitros recentes, de 1255 para 90 opções.

A definição desta nova tela teve como base a experiência do analista-programador e dono do site, além de estar em linha com a identificação da preferência dos usuários pelas informações mais recentes relacionadas com o clube. Foram colocadas perguntas sobre a preferência entre as telas antigas e novas, além de uma avaliação sobre as novas telas e quando necessário foram feitas entrevistas com os usuários que preferiram ou deram avaliações negativas das novas telas.

5.4.4

Melhoria da identidade visual do site

A análise de *logs* identificou que existiam consultas com pouco ou nenhum acesso espalhadas por todo o sistema. Uma das seções mais atingidas por esse pouco acesso era a seção de gráficos. No questionário, perguntamos sobre o acesso aos gráficos: somente 40% dos usuários respondeu que fazia consultas e somente 20% o fazia regularmente. Combinado a isso, a análise de *logs* também identificou que 9 entre 10 das páginas com maiores acessos do sistema eram páginas oriundas de links internos do sistema. Porém, o questionário identificou que mais de 30% dos usuários não sabiam que tais links existiam e mais de 70% julgaram ser muito importante a existência deles.

De acordo com os usuários do sistema, a seção de especiais estava mal formatada e algumas consultas poderiam estar perdidas em subseções que não tinham correlação entre a consulta e a seção em que a consulta estava armazenada, além de ser necessário percorrer um caminho longo para o acesso as consultas.

Alinhando as informações coletadas com os objetivos de maior engajamento, sistema mais interativo, integração das consultas e gráficos, sugerimos melhorias na identidade visual do site:

1. eliminar telas intermediárias com a criação de um sub-menu na tela principal;

2. organizar as consultas especiais em grupos de interesse;
3. para alimentar o sistema de recomendação, criar *tooltips* que informam o que cada sub-seção dos especiais está oferecendo de consulta;
4. criar novos links internos que relacionem à consulta mostrada em texto com o gráfico correspondente; e
5. informar de forma mais clara a existência dos links.

A definição dessas novas telas teve como base a experiência do analista-programador e dono do site. Foram colocadas perguntas sobre a preferência entre as telas antigas e novas, além de uma avaliação sobre as novas telas. Quando necessário, foram feitas entrevistas com os usuários que preferiram ou deram avaliações negativas das novas telas.

5.5

Avaliação das Modernizações Realizadas

Nesta seção apresentamos os resultados obtidos na avaliação das modernizações realizadas no site <https://www.fluzao.xyz> até o momento da redação desta dissertação.

Detalhamos as avaliações do sistema de recomendação Subseção 5.5.1, o redesign das consultas na Subseção 5.5.2, a nova identidade visual do site na Subseção 5.5.3 e os canais de comunicação do site na Subseção 5.5.8, conforme especificado na seção 4.5.

5.5.1

Avaliação do Sistema de Recomendação

Esta avaliação consistiu em coletar dados relativos ao sistema de recomendação implementado com base nas sugestões de modernização e compará-los com dados históricos do uso da página principal antes dos ajustes.

5.5.1.1

Objetivo

O principal objetivo desta avaliação foi verificar se o tempo médio de engajamento dos usuários na tela principal do sistema aumentou e se (e o quanto) os usuários utilizaram os *links* entre as páginas disponibilizadas para a expansão das consultas.

5.5.1.2

Procedimento

A partir da primeira implementação em 27/01/2024, criamos um novo *log* para capturar o tempo médio de engajamento na tela principal do sistema. Extraímos os dados históricos desde 03/08/2023, data da primeira coleta; posteriormente, extraímos os dados semanalmente até o dia 04/08/2024. Geramos gráficos tendo como base o tempo médio de engajamento por data durante todo o período; depois dividimos o dataset em duas partes – a primeira de 03/08/2023 até 26/01/2024 e a segunda a partir de 27/01/2024 –; calculamos as médias totais de engajamento e identificamos a diferença entre elas.

Seguem abaixo os passos realizados.

1. Extrair os *logs* de navegação da página principal do Google analytics, com foco no tempo médio de engajamento.
2. Executar o algoritmo de análise de acesso da página principal e medir o tempo médio de engajamento desde 03/08/2023 até 04/08/2024.
3. Gerar gráficos para apoiar a análise dos resultados.
4. Calcular o tempo médio de engajamento até 27/01/2024.
5. Calcular o tempo médio de engajamento após 27/01/2024.
6. Realizar o cálculo para medir a porcentagem de diferença.

A combinação do tempo médio de engajamento e do resultado das avaliações do questionário dão a medida sobre os benefícios da implantação do sistema de recomendação como tela inicial.

5.5.1.3

Material

Para esta análise, criamos o programa `Analise_Recomendacao.ipnyb`, listado no Apêndice B. As funções do programa são: ler o *log* de engajamento da tela principal, gerar gráficos tendo como base o tempo médio de engajamento por data durante todo o período; dividir o dataset em duas partes – a primeira de 03/08/2023 até 26/01/2024 e a segunda a partir de 27/01/2024 –; calcular as médias totais de engajamento dos dois datasets separadamente e calcular a diferença percentual entre elas.

5.5.1.4

Fonte de Dados

A implementação foi realizada no dia 27/01/2024; têm sido coletados *logs* de dados de todos os usuários que acessaram o site desde então.

Aplicamos um questionário com questões específicas sobre os sistemas de recomendação, iniciado em 01/07/2024 e finalizado em 31/07/2024. Obtivemos 38 respostas, conforme material da Seção A.6.

5.5.1.5

Resultados

O tempo médio de engajamento dos usuários na página principal do sistema aumentou em 250%, conforme visto na Figura 5.10 e Figura 5.11. Anteriormente, a página era estática e proporcionava o acesso a outras áreas do site somente a partir de uma barra de navegação. Atualmente, continua proporcionando o acesso a outras áreas do site e também apresenta para o usuário seis consultas relacionadas ao próximo jogo ou alguma notícia recente importante.

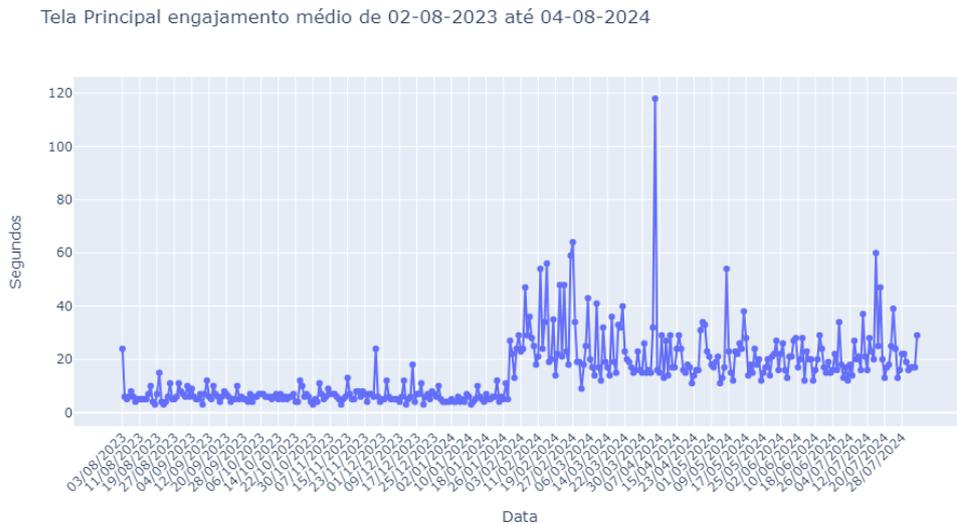


Figura 5.10: Engajamento médio diário

O questionário definido na seção Seção A.6 apresentou um resultado mais apertado: das 38 respostas, 52,6% foram de preferência pela tela nova com o sistema de recomendação. Porém todas as avaliações sobre a tela principal ficaram entre 4 e 7, em uma escala de 1 a 7, crescendo de 1 muito ruim a 7 Excelente, conforme podemos verificar na Figura 5.12 e na Figura 5.13. Com esse resultado, conduzimos entrevistas com alguns dos respondentes para obtermos uma explicação mais detalhada da tela principal e identificar quais os motivos que basearam a escolha de vários pela tela antiga.



Figura 5.11: Engajamento médio mensal

Com relação a tela de entrada do site, qual você prefere?

38 respostas

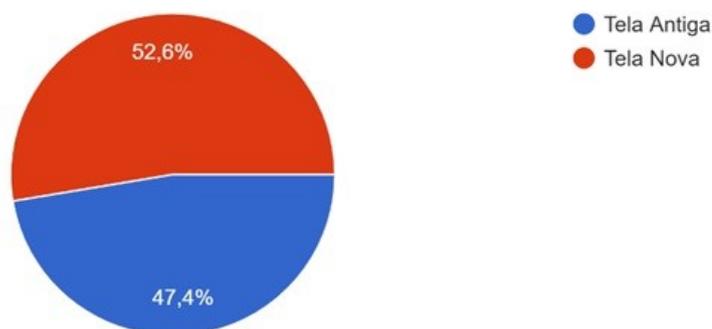


Figura 5.12: Preferência dos respondentes sobre a tela principal

Realizamos entrevistas com alguns dos participantes do questionário que apontaram a preferência para a tela inicial antiga. Dos 18 participantes que indicaram essa preferência no questionário, 12 se prontificaram a participar de uma breve entrevista. Nessa entrevista, utilizamos o seguinte roteiro de tópicos:

- Busca de informações sobre o porquê da escolha pela visualização antiga, neste ponto perguntamos diretamente ao entrevistado o porque ele escolheu a tela antiga como a melhor tela de entrada do sistema;
- Explicação sobre os resultados que levaram à escolha da nova visualização, neste ponto fizemos uma breve explicação de alguns resultados

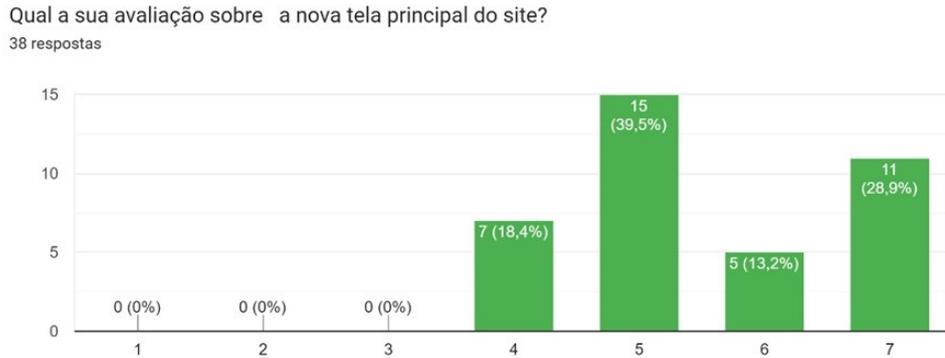


Figura 5.13: Avaliação da tela principal

coletados sobre a preferência de acessos e maior interação do site com o usuário;

- Breve explicação sobre sistemas de recomendação;
- Busca do posicionamento após explicações, neste ponto fizemos uma pergunta sobre após o que foi conversado se o usuário ainda manteria a escolha dele na visualização antiga.

Entrevistamos 8 usuários; destes, 6 informaram que a tela tinha muita informação e poderia causar um pouco de confusão, e somente 2 mudaram a preferência para a tela nova. Porém após as explicações, todos os 8 concordaram com o valor agregado pela nova visualização e se mostraram dispostos a se acostumarem com a nova tela.

Seguindo a mesma linha de avaliação do sistema de recomendação principal, tivemos avaliações de 4 a 7 para 97,4% para os dados relevantes sobre assuntos atuais do clube, conforme a Figura 5.14. Este resultado traz uma boa aprovação para o sistema de recomendação secundário implementado no sistema. A Figura 5.15 e a Figura 5.16 ilustram os sistemas de recomendação implementados no sistema.

5.5.1.6 Execução

Criamos um novo relatório no Google Analytics, com foco no tempo médio de engajamento por tela acessada do sistema. A data de início de coleta foi 03/08/2023 e a data final foi 04/08/2024.

Depois de gerada a extração completa, aplicamos um filtro pela palavra “Principal”, para obtermos os dados apenas de página principal do sistema.

Qual a sua avaliação sobre receber sugestões de consultas relevantes sobre assuntos atuais envolvendo o Fluminense?

38 respostas

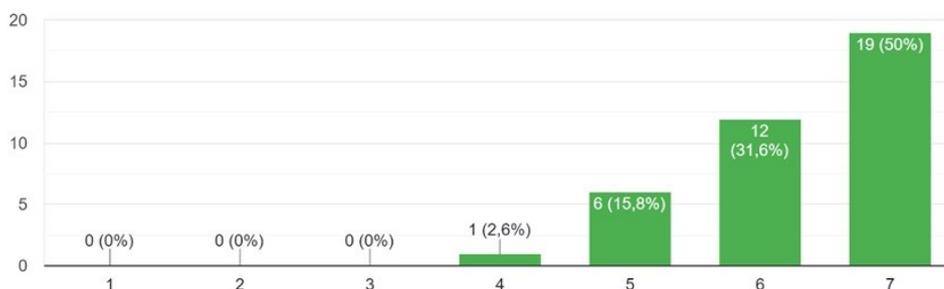


Figura 5.14: Avaliação consultas relevantes na tela principal

Fluminense x Colo Colo-Chile

*Clique aqui e veja mais

	J	V	E	D	GP	GC	Pts	Apr
Mandante	1	1	0	0	2	1	3	100,0%
Visitante	3	1	1	1	4	4	4	44,44%
Total	4	2	1	1	6	5	7	58,33%

160 🔄 8 🔄

Sequências - Libertadores da América

Maior sequência de vitórias (2x) - 4 jogo(s)
 Última vez de 7/2/2012 até 29/3/2012
 Téc - Abel 4 jogo(s)

Maior sequência invicta (1x) - 11 jogo(s) - 6 Vitórias e 5 Empates
 Última vez de 27/6/2023 até 25/4/2024
 Téc - Fernando Diniz 11 jogo(s)

Maior sequência empates (5x) - 2 jogo(s)
 Última vez de 27/6/2023 até 1/8/2023
 Téc - Fernando Diniz 2 jogo(s)

Maior sequência sem vitória (1x) - 9 jogo(s) - 5 Derrotas e 4 Empates
 Última vez de 3/3/1971 até 20/2/2008
 Téc - Nelsinho 6 jogo(s) - Renato Gaúcho 1 jogo(s) - Zagalo 2 jogo(s)

Maior sequência de derrotas (3x) - 2 jogo(s)
 Última vez de 25/5/2023 até 7/6/2023
 Téc - Fernando Diniz 2 jogo(s)

0 🔄 0 🔄

Mais jogos contra Colo Colo

*Clique aqui e veja mais

- 1 - Veludo (1956) - 3 jogos
- 2 - Emilson (1955) - 3 jogos
- 3 - Orlando Pingo Ouro (1953) - 3 jogos
- 4 - Didi (1956) - 3 jogos
- 5 - Pinheiro (1963) - 3 jogos
- 6 - Mário Faria (1951) - 2 jogos
- 7 - Pé de Valsa (1951) - 2 jogos
- 8 - Silas (1951) - 2 jogos
- 9 - Waldir (1951) - 2 jogos
- 10 - Píndaro (1955) - 2 jogos

3 🔄 0 🔄

Evolução contra Colo Colo

*Clique aqui e veja mais

- Jogo 1 - 7/5/1950 0 vit, 0 emp e 1 der (-1) 🕒
- Jogo 2 - 18/5/1950 0 vit, 1 emp e 1 der (-1) 🕒
- Jogo 3 - 7/2/1953 1 vit, 1 emp e 1 der (0) →
- Jogo 4 - 9/4/2024 2 vit, 1 emp e 1 der (1) 🟢

4 🔄 0 🔄

Figura 5.15: Recomendações feitas pelo sistema de recomendação primário

Com esse resultado, exportamos o relatório final somente com os acessos a essa página.



Figura 5.16: Recomendações feitas pelo sistema de recomendação Secundário

5.5.1.7

Análise dos Dados Quantitativos

Nós avaliamos os dados de tempo de engajamento utilizando a data de corte de 27/01/2024 e comparamos os tempos antes e depois. Geramos gráficos com esses dados e calculamos a diferença percentual entre os tempos de engajamento antes e depois do corte para identificamos o crescimento no engajamento do uso da informação, conforme os passos especificados abaixo.

1. Criar um gráfico de linha geral com os tempos médios de engajamento por dia
2. Calcular os tempos médios de engajamento antes e depois de 27/01/2024.
3. Calcular a diferença percentual entre os tempos – gerando um aumento de 30% de engajamento total.

5.5.2

Avaliação do Redesign das Consultas

Nós avaliamos as mudanças realizadas nas consultas dos atributos primários do site: adversário, árbitro, campeonato, estádio, jogador e técnico.

5.5.2.1

Objetivo

Esta avaliação objetivou analisar o tempo médio de engajamento e número de cliques dos usuários nas consultas de atributos e se os conceitos de agilidade e usabilidade implementados foram percebidos e aprovados.

5.5.2.2

Procedimento

Os procedimentos descritos abaixo foram realizados para medir o atingimento de metas para cada um dos parâmetros:

Tempo médio de engajamento

1. Extrair os *logs* de navegação da página principal do Google Analytics com o foco em tempo médio de engajamento.

2. Executar algoritmo de análise de acesso nas páginas de consulta de atributo e medir o tempo médio de engajamento desde 03/08/2023 até a data presente.
3. Gerar gráficos para análise dos resultados.
4. Calcular o tempo médio de engajamento até 24/02/2024.
5. Calcular o tempo médio de engajamento após 24/02/2024.
6. Realizar o cálculo para medir a porcentagem de diferença.

Número de cliques

1. Extrair os *logs* de atividade do Google Analytics com o foco no número de cliques.
2. Executar algoritmo de análise de atividades desde 03/08/2023 até a data presente.
3. Gerar gráficos para análise dos resultados.
4. Calcular o número de cliques médios até 27/01/2024.
5. Calcular o número de cliques médios após 27/01/2024.
6. Realizar o cálculo para medir a porcentagem de diferença.

Aceitação do usuário

1. Aplicar do questionário para avaliar as mudanças.
2. Compilar dos resultados do questionário.
3. Conduzir entrevistas para análise mais minuciosa das alterações, se necessário.
4. Compilar os resultados finais.

5.5.2.3

Material

Para a etapa de avaliação do redesign das consultas, criamos o programa `Analise_Consultas_Primarys.ipnyb`, listado no Apêndice B.

O programa de análise de consultas primárias gera um gráfico de engajamento médio nas telas de consulta aos atributos primários (adversários, árbitros, campeonatos, estádio, jogadores e técnicos); divide o arquivo em duas partes, utilizando Jan/2024 como data de corte; calcula as médias de tempo dos dois períodos e calcula a variação percentual.

No questionário de avaliação aplicado de 01/07/2024 até 31/07/2024, foi feita uma pergunta e uma avaliação sobre a nova tela criada.

5.5.2.4

Fonte de Dados

As implementações vêm sendo realizadas desde o dia 24/02/2024. Vêm sendo coletados dados de todos os usuários que acessaram as consultas de atributos desde então.

Aplicamos um questionário com questões específicas sobre o redesign das consultas dos atributos primários, iniciado em 01/07/2024 e finalizado em 31/07/2024. Obtivemos 38 respostas, conforme material da Seção A.6.

5.5.2.5

Resultados

Os dados coletados no estudo foram contrastados da seguinte maneira: tempos médios de engajamento anteriores x tempos médios de engajamento atuais. Era esperado que o tempo de engajamento aumentasse e que a quantidade de cliques no sistema diminuiria com as novas opções de design.

Com relação ao número de cliques, verificamos uma diminuição de 20% a partir do mês de fevereiro. Os resultados indicam uma tendência de diminuição devido à alteração das telas de consulta de atributos, conforme pode ser visto na Figura 5.17 e na Figura 5.18.

O novo questionário trouxe um resultado bastante favorável aos ajustes realizados. A nova tela de consulta de árbitros foi a preferência de 73,7% dos respondentes, com avaliação acima de 4 para 89,5% dos respondentes, conforme a Figura 5.19 e a Figura 5.20. A tela de consulta de árbitros tinha a maior probabilidade de apresentar dificuldades na busca pelos nomes devido ao *combobox* com 1255 nomes. A nova tela com os ajustes realizados pode ser visualizada na Figura 5.21.

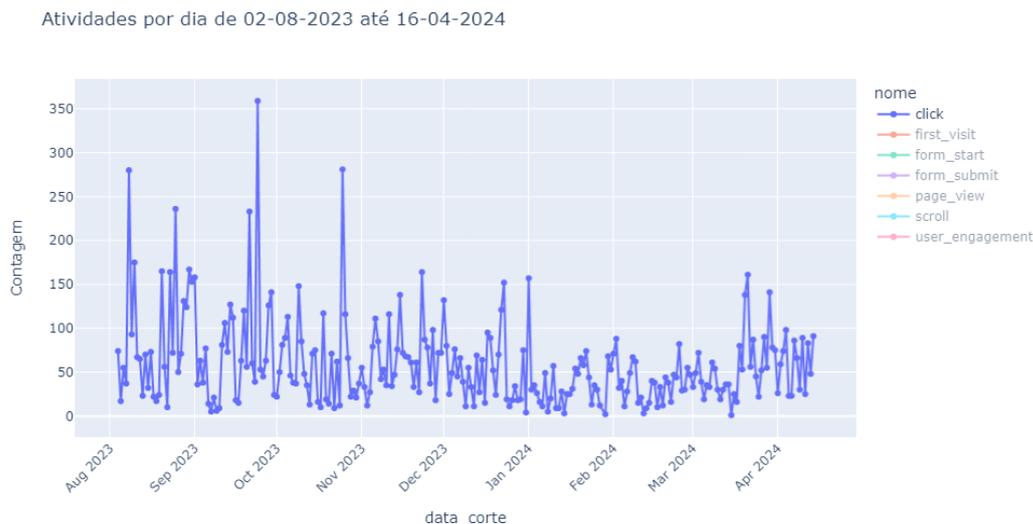


Figura 5.17: Número total de cliques por dia



Figura 5.18: Número total de cliques por mês

5.5.2.6 Execução

Criamos um novo relatório no Google Analytics, que tem foco o tempo médio de engajamento por tela acessada do sistema. A data de início de coleta é 03/08/2023 e a data final é a data corrente - 1.

Depois de gerada a extração completa, aplicamos os filtros pelas principais seções do site: “Adversário”, “Árbitro”, “Estádio”, “Campeonato”, “Jogador” e “Técnico”. Com esse resultado, exportamos o relatório final com os acessos das páginas de consulta de atributo do sistema.

Para a coleta da quantidade de cliques realizados, buscamos o relatório

Qual tela de consulta de árbitros você prefere?

38 respostas

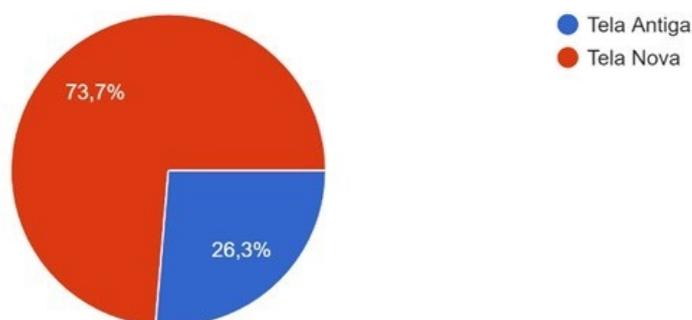


Figura 5.19: Preferência entre telas de consulta

Qual a sua avaliação sobre a divisão em Recentes e Antigos nas consultas de adversário, árbitro, campeonato, estádio, jogador e técnico?

38 respostas

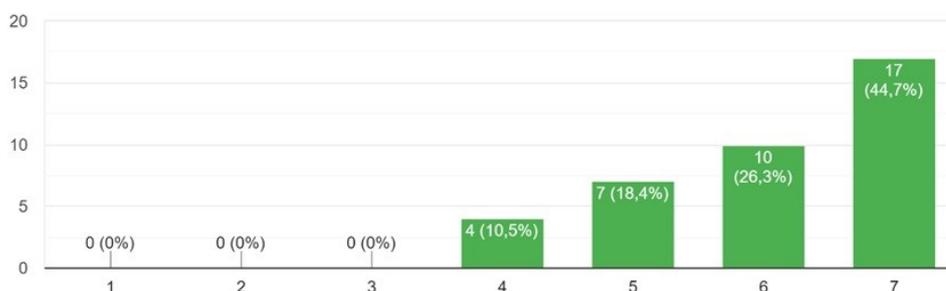


Figura 5.20: Avaliação da nova tela de consulta

de atividades do Google Analytics, isolamos o número de cliques e realizamos as análises dos dados por dia e por mês.

No questionário aplicado, foram incluídas questões sobre a preferência entre a forma antiga e a nova de consulta e foram realizadas entrevistas com os usuários que tiveram posições muito dissonantes do resultado geral.

5.5.2.7

Análise dos Dados Quantitativos

Verificamos o tempo médio de engajamento, utilizando a data de corte de 24/02/2024 e comparando os tempos antes e depois. Da mesma forma, coletamos os dados de quantidade de cliques para a realização das consultas com a mesma data de corte, comparando o antes e o depois. Por fim, aplicamos o



Figura 5.21: Nova tela de consulta

questionário e entrevistas com perguntas relacionadas à preferência do usuário entre a tela antiga e a tela nova de consulta.

Tempo de Engajamento

1. Criamos um gráfico de linha com os tempos médios de engajamento por dia
2. Calculamos os tempos médios de engajamento antes e depois de 24/02/2024 por consulta.
3. Calculamos a diferença percentual entre os tempos de engajamento por consulta - não observamos variações de tempo de engajamento.

Quantidade de cliques

1. Criamos um gráfico de linha com a quantidade de cliques totais do sistema por dia
2. Calculamos o número de cliques antes e depois de 24/02/2024 por consulta.
3. Calculamos a diferença percentual entre as quantidade de cliques – observamos uma pequena diminuição da quantidade de cliques.

Preferência entre telas

1. Criamos questionário com perguntas sobre a preferência entre as telas antigas e novas.
2. Compilamos os resultados do questionário.

3. Realizamos entrevistas com os usuários com respostas mais dissonantes.
4. Compilamos os resultados das entrevistas com preferência em média de 75% pelas novas telas.

5.5.3

Avaliação da Nova Identidade Visual do Site

Nesta seção relatamos a avaliação das alterações realizadas em aspectos visuais e de navegação do site para melhorar a experiência do usuário. Analisamos os resultados da nova seção de especiais desde seu acesso até a sua formatação e agrupamento de consultas, bem como as novas telas de visualização dos atributos primários, atributos combinados, rankings e listas.

5.5.3.1

Objetivo

O objetivo desta avaliação foi verificar se a usabilidade do sistema e a experiência do usuário na seção especiais e nas telas de visualização melhoraram com os ajustes realizados através das sugestões feitas pelos dados coletados através da abordagem.

5.5.3.2

Procedimento

Com base nos resultados das primeiras análises de *logs* referentes aos números de acesso às telas das consultas especiais e à utilização dos links internos e consultas mais/menos acessadas, desenhamos sugestões para tornar o acesso às consultas mais simples e fácil, com uma melhor organização das consultas especiais. Isso tornou as visualizações de dados mais modernas, com mais destaque para os links e inclusão de conteúdos visuais que ajudam a identificar o que foi consultado.

O procedimento criado para avaliação desses ajustes foi concentrado na aplicação do questionário que foi aplicado no mês de Julho/2024.

Segue o passo a passo deste procedimento:

1. Criar o questionário com questões relativas a nova identidade visual do sistema.
2. Coletar as respostas.
3. Compilar os resultados do questionário.

4. Criar o roteiro das entrevistas, se necessário.
5. Realizar as entrevistas, se necessário.
6. Compilar os resultados.

5.5.3.3

Material

Elaboramos um questionário, que foi aplicado entre os dias 01/07/2024 e 31/07/2024 (ver Seção A.6). Fizemos diversas perguntas para avaliar a nova identidade visual do site. A seguir detalhamos os resultados.

5.5.3.4

Tela principal de visualização do jogo - Resultado

A tela de visualização do jogo foi uma das poucas telas que não foi substituída. As duas visualizações permanecem ativas dentro do sistema. Essa escolha se mostrou acertada, tendo em vista os resultados do questionário. Conforme a Figura 5.22, 81,6% dos participantes preferem a visualização antiga. Dessa forma, a visualização antiga, Figura 5.24, voltou a ser a principal e a visualização nova, Figura 5.25, passou a ser uma forma alternativa de visualizar os dados. Mesmo não sendo a preferida, 31,6% avaliaram a nova tela de forma neutra e 55,2% de forma positiva, conforme a Figura 5.23.

Qual a tela de exibição do jogo você prefere?

38 respostas

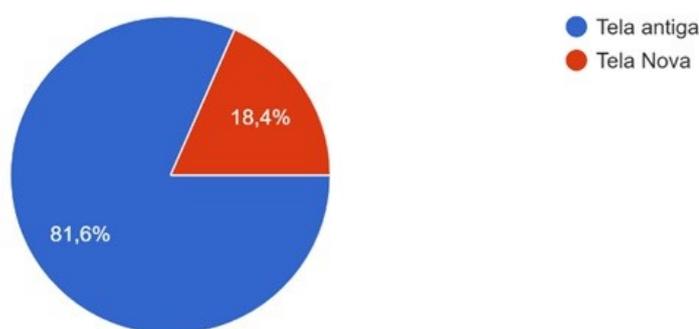


Figura 5.22: Preferência sobre a tela de jogo

Qual a sua avaliação sobre o novo design da tela de jogos?

38 respostas

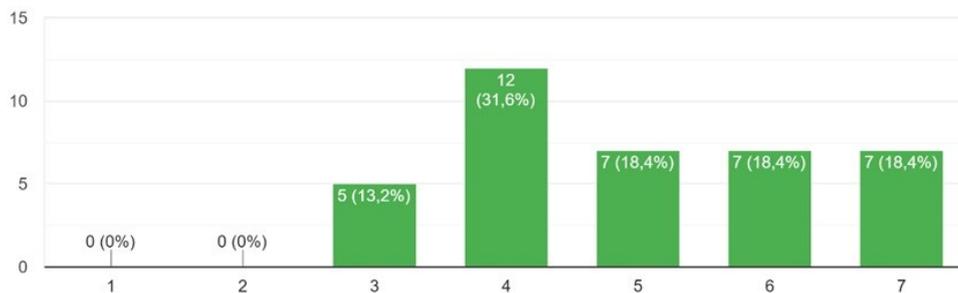


Figura 5.23: Avaliação da nova tela de jogo



Figura 5.24: Tela de jogo antiga

5.5.3.5

Organização da seção dos Últimos 12 Jogos - Resultados

Conforme a Figura 5.26 e a Figura 5.27, 89,5% preferiram a nova forma de visualização e somente 5,3% avaliaram de forma negativa a mudança. Conduzimos entrevistas com esses dois respondentes e as respostas foram na



Figura 5.25: Tela de jogo nova

linha de serem um pouco mais resistentes a mudanças. Porém, com o tempo, podem se acostumar e isso não seria um obstáculo para continuarem usando o site e suas informações. A nova visualização pode ser vista na Figura 5.28. O ajuste realizado teve foco na melhoria de visualização dos links dentro do sistema.

Qual organização da seção dos últimos 12 jogos você prefere?
38 respostas

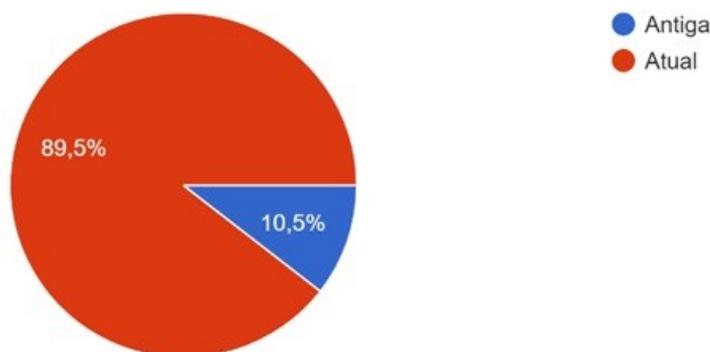


Figura 5.26: Preferência do novo formato dos últimos 12 jogos

5.5.3.6

Telas de apresentação dos atributos primários - Resultados

Parte da modernização do sistema teve foco de melhorar a visualização e facilitar a utilização dos *links*, mantendo os dados do adversário mais agrupados no topo da tela, conforme a Figura 5.31. Tivemos um resultado

Qual a sua avaliação sobre a nova organização dos últimos 12 jogos?

38 respostas

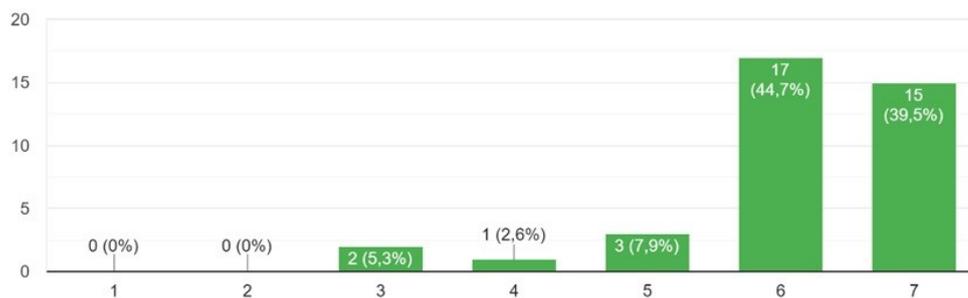


Figura 5.27: Avaliação do novo formato dos últimos 12 jogos

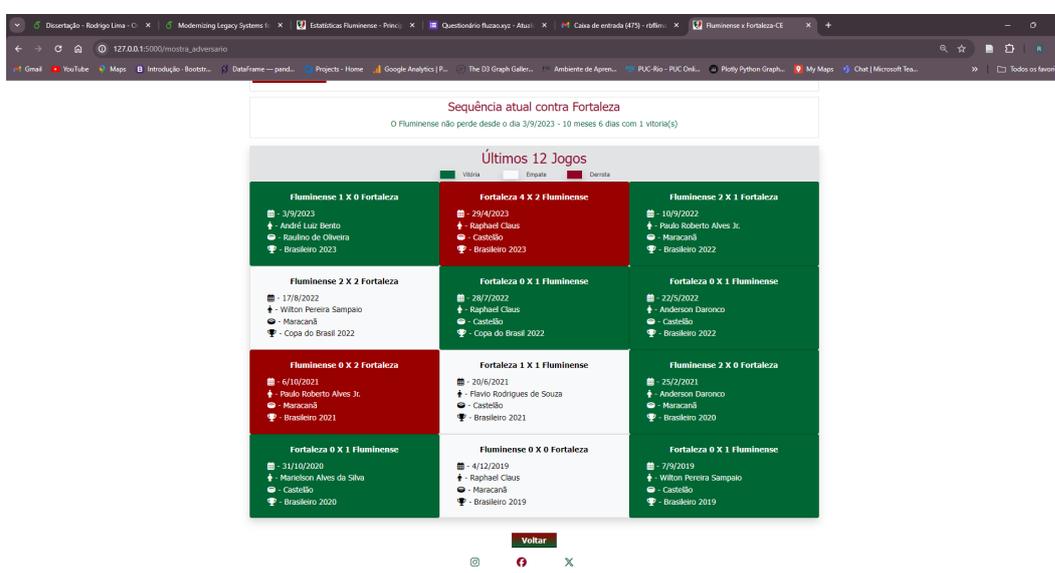


Figura 5.28: Novo formato dos últimos 12 jogos

muito satisfatório, com mais de 86,8% de preferência pela nova tela, conforme visto na Figura 5.30, com aprovação de 97,4%. Conforme a Figura 5.29, foi recebida uma avaliação de grau 1 para a nova visualização, mas o usuário a escolheu como a preferida. Ao entrarmos em contato com esse usuário, recebemos a informação que ele errou ao informar sua avaliação numérica.

5.5.3.7

Forma de acesso à seção de especiais - Resultados

A seção de especiais foi substituída de uma tela composta de uma série de botões para um submenu na tela principal, conforme visto na Figura 5.32 e na Figura 5.33. Além disso, alteramos os conteúdos de cada seção, tornando-os mais organizados e com explicações sobre o que pode ser encontrado naquela seção. Os resultados foram contraditórios: 60,5% preferiram a visualização

Qual a sua avaliação sobre a nova tela de visualização dos dados de adversário?

38 respostas

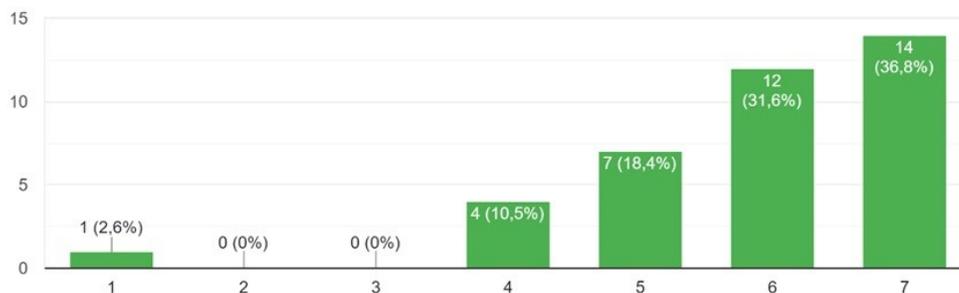


Figura 5.29: Preferência da apresentação dos atributos primários

Qual tela de visualização dos dados do adversário você prefere?

38 respostas

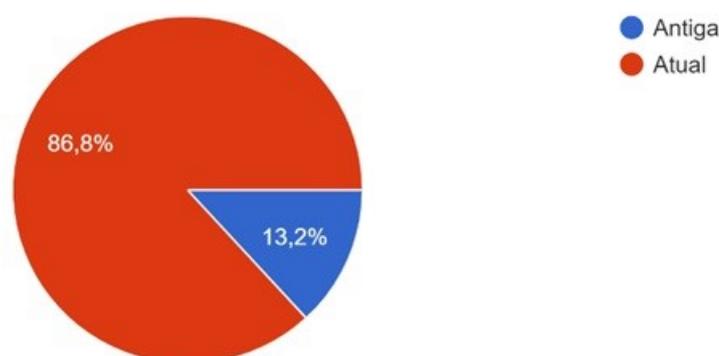


Figura 5.30: Avaliação da nova apresentação dos atributos primários

antiga, mesmo que a avaliação da nova visualização tenha sido aprovada por 89,7% dos respondentes. Já a distribuição das consultas foi aprovada por 100% dos respondentes, conforme visto na Figura 5.34, na Figura 5.35 e na Figura 5.36. Um exemplo das novas telas de especiais pode ser vista na Figura 5.37.

5.5.3.8

Nova visualização de rankings - Resultados

A nova visualização dos rankings de jogadores e adversários trouxe mais aspectos visuais, conforme pode ser visto na Figura 5.38 e na Figura 5.39. Em ambos os casos, a preferência ficou com a nova forma de visualização, com 97,4% para o ranking de jogadores e 94,7% para o ranking de adversários. A avaliação do ranking de jogadores teve 1 nota 3, enquanto o ranking de



Figura 5.31: Nova tela de apresentação dos adversários

adversários todas as notas foram no mínimo 4, conforme pode ser visto na Figura 5.40, na Figura 5.41, na Figura 5.42 e na Figura 5.43.

5.5.3.9

Nova visualização de consultas combinadas – Resultados

A nova visualização de parâmetros combinados trouxe mais aspectos visuais, conforme pode ser visto na Figura 5.44. A preferência ficou com a nova forma de visualização, com 89,5% das respostas. A avaliação da nova tela teve apenas uma nota 3, conforme pode ser visto na Figura 5.45 e na Figura 5.46.

5.5.4

Participantes

Participaram do estudo 38 pessoas de diversos perfis, jornalistas, torcedores e historiadores de futebol.

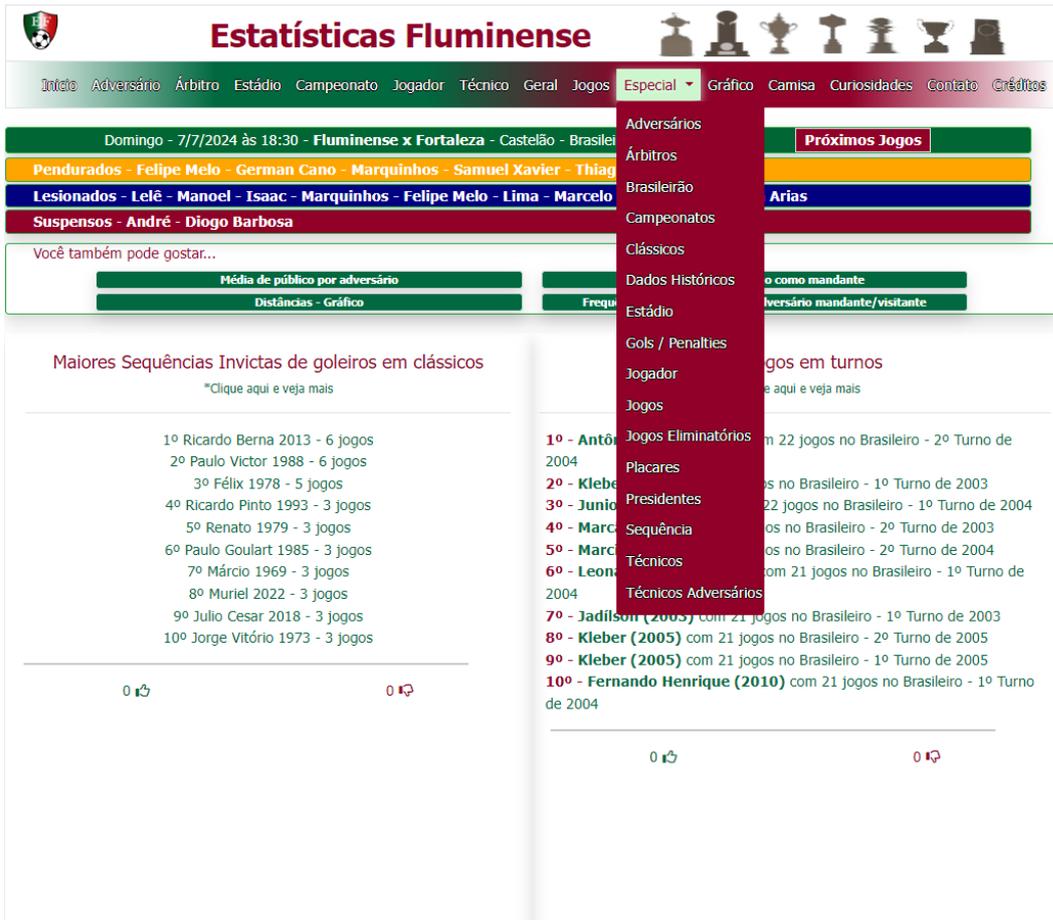


Figura 5.32: Novo formato de acesso à seção de especiais



Figura 5.33: Seção de especiais antiga

5.5.5 Execução

Foi aplicado um questionário com questões específicas sobre o redesign das consultas dos atributos primários, iniciado em 01/07/2024 e finalizado em

Na sua opinião qual das duas formas de acesso a seção de especiais é melhor?
38 respostas

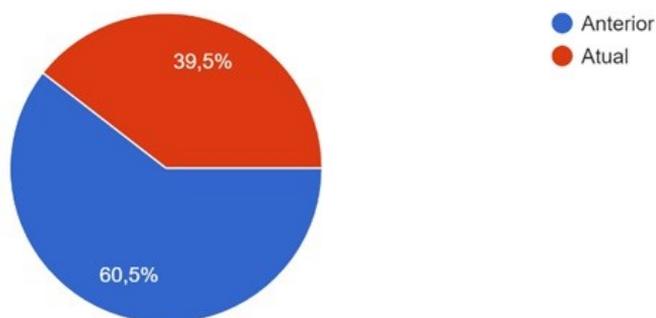


Figura 5.34: Preferência da apresentação da seção especiais

Qual a sua avaliação sobre a troca da seção Especiais do site por um submenu na tela principal?
38 respostas

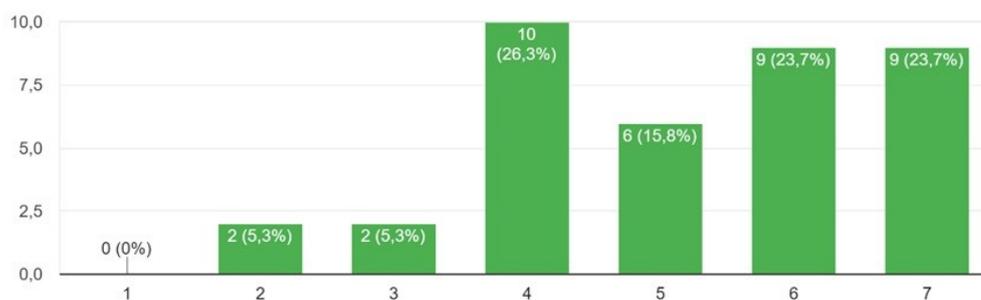


Figura 5.35: Avaliação da nova apresentação da seção especiais

Qual a sua avaliação sobre a nova formatação das consultas na seção especiais?
38 respostas

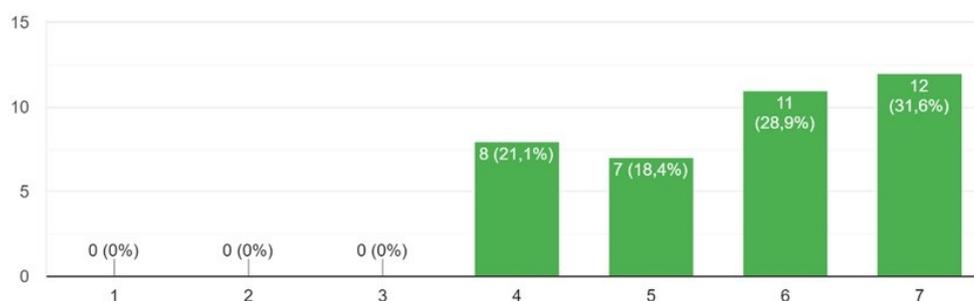


Figura 5.36: Avaliação da distribuição das consultas da seção especiais

31/07/2024. Obtivemos 38 respostas, conforme material da Seção A.6.



Figura 5.37: Nova distribuição da seção Especiais

5.5.6

Análise dos Dados Quantitativos

Utilizamos os gráficos do Google Analytics em conjunto com as respostas dos questionários e entrevistas, capturando o grau de satisfação dos usuários com a nova identidade visual do site proposta após as sugestões de melhoria. O procedimento seguido para a análise foi o seguinte:

1. Utilizar os gráficos fornecidos pela ferramenta Google Forms.
2. Compilar as respostas dos questionários analisando os usuários que avaliaram negativamente as mudanças.
3. Realizar entrevistas com os usuários a fim de entender as razões que levaram a avaliação negativa.

A nova identidade visual do site foi aprovada por mais de 85% dos respondentes. Os resultados de todos os diferentes questionamentos podem ser analisados nas respostas do questionário de avaliação na seção Seção A.6

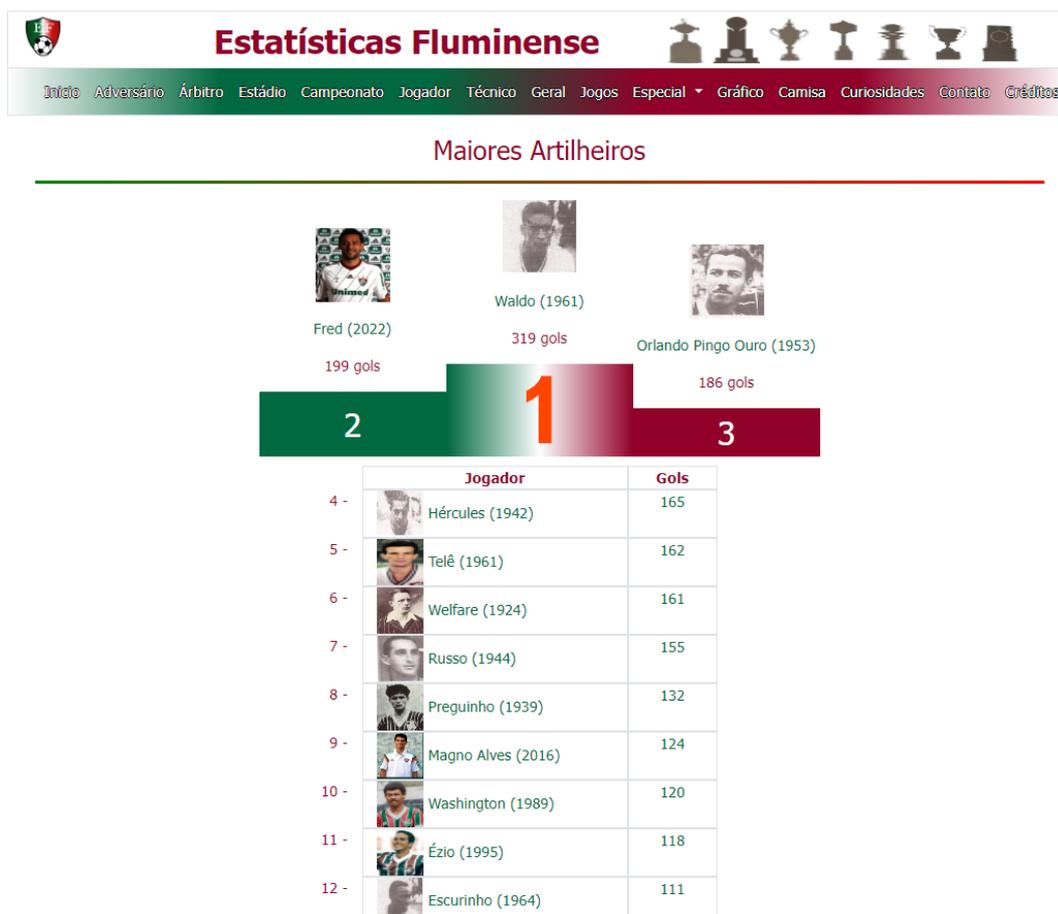


Figura 5.38: Novo ranking de jogadores

5.5.7

Discussão

Conforme descrito na seção 5.5.6, tivemos um resultado positivo sobre as alterações realizadas, atingindo o objetivo de melhorar a experiência na navegação do site. O ajuste da tela de jogo que não atingiu o objetivo foi o único ajuste em que uma nova tela foi adicionada e não substituiu a tela antiga. Desta forma, é possível retornar à tela antiga como a principal e a tela nova fica como uma segunda visualização, mantendo a experiência preferencial do usuário no acesso ao sistema.

5.5.8

Avaliação dos Canais de Comunicação do Site

Após a criação de um canal de comunicação do site nas três maiores redes sociais da atualidade (X⁶, Instagram⁷ e Facebook⁸), buscamos identificar o impacto desse movimento no número de acessos e engajamento no site.

⁶<https://twitter.com/EstatisticasFlu>

⁷<https://www.instagram.com/estatisticasfluminense>

⁸<https://www.facebook.com/estatisticas.fluminense>



Figura 5.39: Novo ranking de adversários

Qual tela de ranking de artilheiros você prefere?

38 respostas

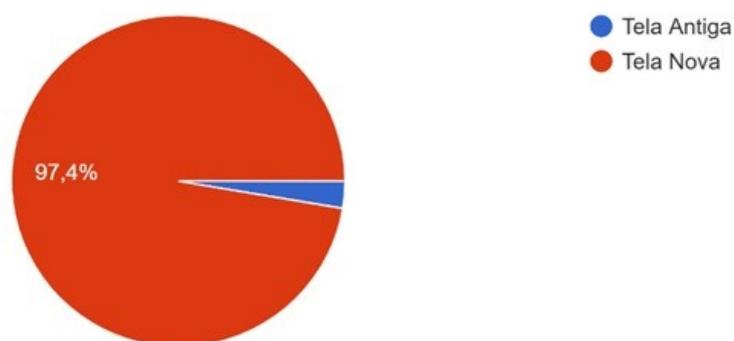


Figura 5.40: Preferência da apresentação de ranking de jogadores

Qual a sua avaliação sobre o novo design dos rankings existentes no site?

38 respostas

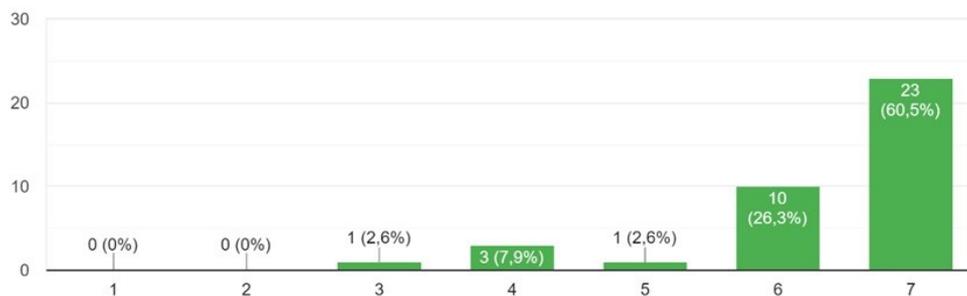


Figura 5.41: Avaliação da nova apresentação de ranking de jogadores

Qual das telas de rankings de adversários você prefere?

38 respostas

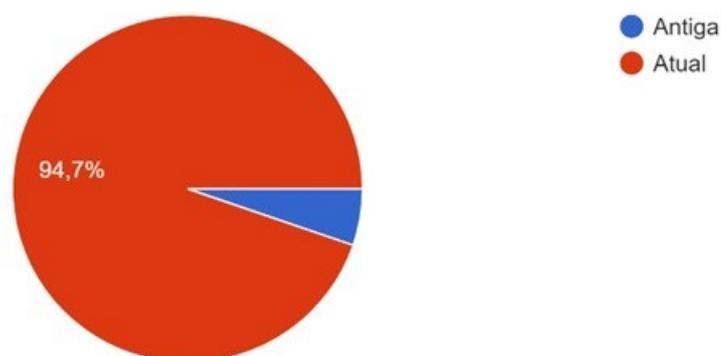


Figura 5.42: Preferência da apresentação de ranking de adversários

Qual a sua avaliação sobre as novas telas de ranking de adversários?

38 respostas

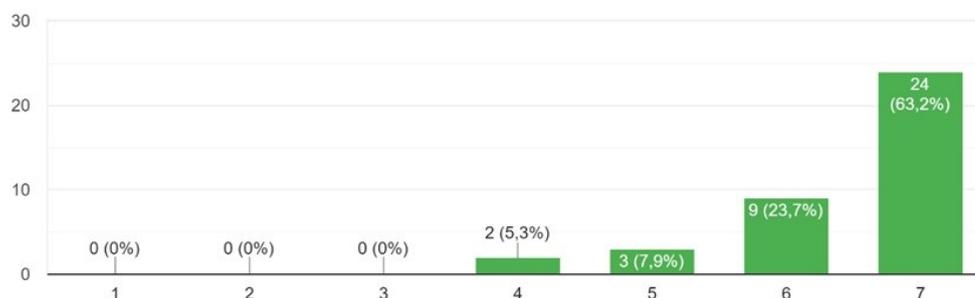


Figura 5.43: Avaliação da nova apresentação de ranking de adversários



Figura 5.44: Nova tela de visualização de especiais

Qual das telas de informações combinadas você prefere?

38 respostas

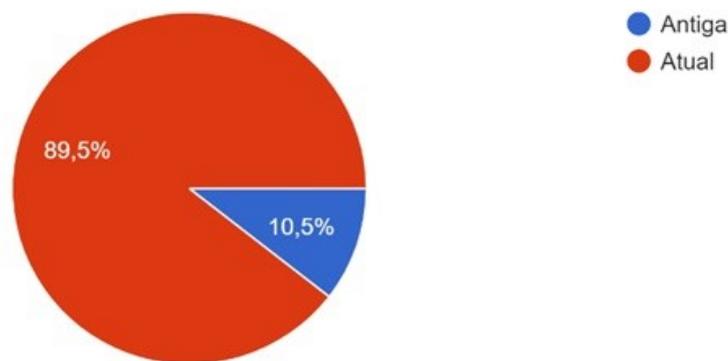


Figura 5.45: Preferência da apresentação da visualização de especiais

5.5.8.1 Objetivo

O objetivo desta avaliação foi verificar se o número de acessos e a origem dos acessos sofreu algum impacto após a criação do canal de comunicação via redes

Qual a sua avaliação sobre as novas telas de informações combinadas?

38 respostas

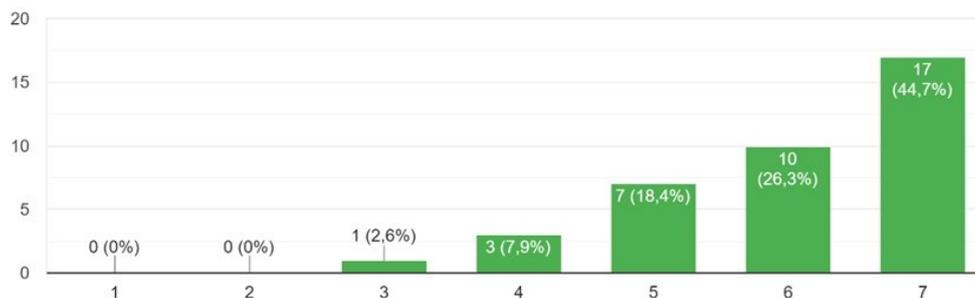


Figura 5.46: Avaliação da nova apresentação da visualização de especiais sociais, além da qualidade das postagens realizadas.

5.5.8.2

Procedimento

Analisamos o número de acessos ao sistema que se originaram em uma das três redes sociais através de alguma das postagens realizadas no período de 18/01/2024 até 02/08/2024. Além disso, analisamos os números obtidos por cada postagem realizada com informações extraídas do site e sempre relacionadas com as principais notícias do momento. Para isso, seguimos estes passos:

1. Extrair os *logs* de navegação da página principal do Google Analytics com o foco em número de usuários e origem do acesso.
2. Executar algoritmo de análise de acesso nas páginas de consulta de atributo e medir o número e origem dos acessos desde 03/08/2023 até 04/08/2024.
3. Gerar de gráficos para análise dos resultados.
4. Verificar o número de acessos e origem de acessos até 17/01/2024.
5. Verificar o número de acessos e origem de acessos após 17/01/2024.
6. Comparar os resultados – o número de acessos via rede social não apresentou nenhuma mudança em relação ao que vinha acontecendo. Portanto, não foi possível avaliar se o aumento de acessos teve alguma relação com o novo canal de comunicação.

5.5.8.3

Material

Para avaliar o impacto dos diversos canais de comunicação do site, elaboramos o programa `Analise_acessos.ipnyb`, listado no Apêndice B

Esse programa busca os dados da origem dos acessos, gera gráficos das quantidades por origem/data, separa os datasets com data de corte em 16/02/2024 e compara os resultados. Como não tivemos variação nos dados, não foi necessário gerar a variação percentual.

5.5.8.4

Participantes

Participaram do estudo 38 pessoas de diversos perfis, jornalistas, torcedores, historiadores de futebol, além de todos os novos 1190 seguidores dos perfis do site nas três plataformas.

5.5.8.5

Execução

Analizamos os *logs* de acesso ao sistema relacionados a origem do acesso, de onde o usuário fez a entrada inicial no sistema, comparando os acessos antes e depois do dia 19/01/2024. Analizamos também os dados fornecidos pela rede social X (twitter) em cada postagem feita, incluindo o número de *links* clicados, o total de engajamento e os números de curtidas e repostagens realizadas.

Aplicamos um questionário com questões específicas sobre o redesign das consultas dos atributos primários, iniciado em 01/07/2024 e finalizado em 31/07/2024. Obtivemos 38 respostas, conforme material da Seção A.6.

5.5.8.6

Análise dos Dados Quantitativos

Coletamos os dados de origem de acesso do Google Analytics, os dados relacionados a cada postagem realizada na rede social X (twitter), colocamos perguntas relacionadas às postagens no questionário aplicado e compilamos os resultados, conforme os passos a seguir.

1. Coletar dados de origem de acesso no Google Analytics
2. Coletar os dados de cada postagem na rede social X
3. Coletar os dados dos questionários nas perguntas relacionadas com as postagens e redes sociais
4. Compilar as respostas dos questionários

5. Realizar entrevistas com os usuários a fim de entender as razões que levaram a avaliação negativa, se necessário.

5.5.8.7

Resultados

A entrada nas redes sociais não proporcionou nenhum impacto no número de acessos ao site e nem na sua originação. Ao todo, o perfil do site tem 822 seguidores no X. Nas outras redes sociais, não alcançamos números impactantes de seguidores. Foram realizadas mais de 100 postagens em todas as redes sociais de fevereiro de 2024 a setembro de 2024. O número de acessos ao site teve um leve aumento, porém as entradas no site via rede social também se mantiveram estáveis. De todas as postagens feitas na rede social X, tivemos menos de 190 cliques no link que apontava para a página principal do site, desta forma não conseguimos relacionar novos canais de comunicação com aumento de acessos ao site.

Sobre as redes sociais, 52,7% dos usuários que responderam a pesquisa no site disseram que não visualizam ou que pouco visualizam as postagens realizadas, conforme a Figura 5.47. Com relação a gostar do conteúdo proposto, 47,4% informaram curtir o conteúdo, 18,4% ficaram com a posição neutra e 26,3% não possuíam redes sociais, conforme demonstrado na Figura 5.48.

Foram criados os canais de comunicação do site nas principais redes sociais, você costuma visualizar o conteúdo das postagens?

38 respostas



Figura 5.47: Visualização postagem rede social X

5.5.8.8

Discussão

Segundo a análise dos *logs*, a criação do canal de comunicação nas redes sociais não surtiu o efeito esperado, embora o resultado do questionário tivesse



Figura 5.48: Avaliação do conteúdo postado na rede social X

indicado que a existência deste canal de comunicação seria uma boa iniciativa. Alguns fatores podem ter influenciado esse resultado. O primeiro é a fase atual vivida pelo clube, que influencia diretamente a vontade do usuário de consumir as informações. Outro ponto que pode ter influência, porém não é objetivo desse estudo se aprofundar, é o comportamento do usuário de rede social: ele pode receber a informação na postagem e isso bastar, e desta forma ele não vai acessar o site em busca de mais dados sobre a postagem ou até mesmo mais informações sobre o assunto.

5.5.9

Considerações sobre as Avaliações

A abordagem de modernização de sistemas legados baseada na análise de *logs*, questionários e entrevistas trouxe informações importantes sobre os padrões de navegação dos usuários e se mostrou ser um bom suporte para elaborar recomendações de modernização. A grande maioria das recomendações foi aprovada e bem avaliada pelos usuários. Com isso, o sistema agora proporciona uma melhor experiência para o usuário, com informações mais acessíveis em todos os níveis e um futuro ainda mais promissor quando todos os ajustes tiverem sido implementados. Além disso, melhora ainda mais a documentação do sistema, facilitando futuras análises e modernizações, tudo isso mantendo as funcionalidades do sistema com baixo custo e baixo esforço.

6

Avaliação da Abordagem por Analistas Externos

Este capítulo descreve como avaliamos a utilização da nossa proposta por terceiros, apresenta os resultados da avaliação e discute suas implicações para a abordagem de modernização de sistemas legados com foco na análise de *logs*.

6.1

Objetivo

O objetivo desta avaliação foi avaliar a abordagem e sua documentação, realizando um estudo com analistas externos, ou seja, pessoas que não estiverem envolvidas na definição nem na aplicação da abordagem.

6.2

Participantes

A avaliação foi realizada com três participantes: um aluno do primeiro período de Ciência da Computação e dois profissionais com larga experiência na análise de sistemas e criação e manutenção de sistemas. Os dois profissionais trabalham especialmente com Cobol (para UNIX), interagindo com Java, ColdFusion, Angular e REACT. Participam da concepção e codificação de interfaces para corretores, pessoal interno da empresa e segurados, com foco na integração com os sistemas legados de uma grande empresa de seguros. Ambos possuem mais de 20 anos de experiência na área, mas sem experiência na linguagem Python, que foi utilizada para a criação dos programas de apoio à abordagem.

6.3

Material

Para a avaliação com os analistas externos, usamos um documento de definição da abordagem (Seção A.7), os programas utilizados na análise de *logs* (Apêndice B), que estão armazenados no Google Colab, conforme seção Seção A.9, e um roteiro de entrevista (Seção A.8).

6.4

Procedimento

Para avaliar a proposta, seguimos os passos a seguir.

1. Conduzimos com cada participante uma sessão de apresentação da abordagem, o objeto de estudo e os seus desafios, em que apresentamos os programas criados para a análise de *logs*. Para evitar influenciar sua análise, geramos novos notebooks limpos para que fosse possível realizar as análises sem interferência de uma análise prévia.
2. Disponibilizamos as ferramentas para que o participante pudesse analisar a abordagem.
3. Respondemos aos questionamentos para sanear qualquer tipo de dúvida que surgisse.
4. Fizemos uma entrevista com o participante para identificar os pontos fortes e os pontos fracos da abordagem proposta.

6.5

Discussão dos resultados

Relatamos a seguir o resultado das sessões de estudo conduzidas com os participantes voluntários para análise da abordagem.

6.5.1

Participante 1 - Aluno do primeiro período

O estudo ocorreu nos dias 03/07/2024 e 10/07/2024. Na primeira sessão, apresentamos o objeto de estudo e sua documentação, fizemos uma apresentação sobre a abordagem e, como usuário-chave, descrevemos alguns dos desafios. Após essa introdução, apresentamos os programas de análise de *logs*, conforme Seção A.9, o Google Analytics¹ e seus relatórios e os arquivos de entrada já formatados. Solicitamos que o participante fizesse uso da abordagem, executando os programas de análise de *logs* e focando em dados de 03/08/2023 até 31/12/2023. O objetivo era propor algum tipo de sugestão; as dúvidas deveriam ser enviadas e seriam esclarecidas de imediato.

Durante a segunda sessão, o participante preencheu ao formulário descrito na Seção A.7, e conversou sobre o que achou sobre cada fase da abordagem, o que acha que pode melhorar e os pontos fortes e fracos.

Durante a segunda sessão, o participante expressou algumas dúvidas sobre a lógica do sistema para análise de *logs*. Esclarecemos as dúvidas lógicas que surgiram e o participante conseguiu finalizar o experimento, ficou evidenciado que o sistema para análise de *logs* contém algumas especificidades que trazem dificuldades para profissionais com pouca experiência.

¹<https://analytics.google.com>

Nos itens abaixo relatamos breves observações sobre o andamento fase a fase, perguntas, dificuldades, impressão sobre o código, pontos fortes, pontos fracos, sugestões e avaliação final da abordagem.

Resultados:

- **Fase 1** - A definição da ferramenta e busca dos *logs* foi percebida como fácil de entender e executar, porém o entendimento sobre o que tem que ser analisado e como analisar somente através dos programas foi bem mais difícil.
- **Fase 2** - Tendo sido quebrado o obstáculo da primeira análise, elaborar o questionário e o roteiro de entrevista não foram um desafio; foi mais tranquilo de fazer.
- **Fase 3** - Compilar os resultados das análise de *logs*, questionários e entrevistas não é uma tarefa complicada; a tarefa de propor sugestões e iniciar o desenvolvimento é subjetiva, porém possível de se resolver. No caso do desenvolvimento, podemos encontrar problemas técnicos no caso de não estarmos muito familiarizados com a linguagem de programação escolhida.
- **Fase 4** - O participante não enxergou dificuldade em trabalhar com os requisitos da fase 4: uma vez que a análise de *logs* inicial foi feita, o trabalho vira um incremento e fica simples de realizar.
- **Perguntas** - Sobre a abordagem não tiveram perguntas, somente questionamentos técnicos sobre os programas de análise de *logs* e sua lógica.
- **Dificuldades** - Análise de *logs* inicial.
- **Impressões sobre códigos** - Os programas ficaram muito específicos e podem complicar quando tiverem que ser utilizados em outro sistema. Com as necessidades de análise deste sistema, eles também podem produzir algum tipo de viés na análise.
- **Mais gostou** - Como a abordagem conecta as análises iniciais e ajustes ao desenvolvimento dos questionários.
- **Menos gostou** - Busca inicial dos *logs*, dificuldade para entender e interpretar os programas de análise.
- **Sugestões de melhoria** - Devido à falta de experiência, foi difícil sugerir mudanças na abordagem, porém a informação sobre a dificuldade de interpretar as lógicas dos programas trouxe um dado valioso para uma melhoria, voltada para profissionais sem experiência.
- **Avaliação Final** - Considera que é possível utilizar a abordagem de forma independente.

6.5.2

Participantes 2 e 3 - Profissionais com larga experiência na análise de sistemas e criação e manutenção de sistemas, especialmente em Cobol.

O estudo ocorreu nos dias 15/07/2024 e 22/07/2024, via vídeo-conferência com os 2 profissionais participando ao mesmo tempo, porém dando suas respostas de forma individual. Na primeira sessão, apresentamos o objeto de estudo e sua documentação, fizemos uma apresentação sobre a abordagem e, como usuário-chave, descrevemos alguns dos desafios. Após essa introdução, apresentamos os programas de análise de *logs*, o Google Analytics e seus relatórios e os arquivos de entrada já formatados. Foi solicitado que o participante fizesse uso da abordagem realizando execuções dos programas de análise de *logs*, focando em dados de 03/08/2023 até 31/12/2023 visando que fosse possível atingir o objetivo de propor algum tipo de sugestão; as dúvidas deveriam ser enviadas e seriam esclarecidas de imediato.

Durante a segunda sessão, os participantes preencheram ao formulário descrito na Seção A.7, e conversamos sobre o que acharam sobre cada fase da abordagem, o que acharam que pode melhorar e os pontos fortes e fracos.

Resultados:

- **Fase 1** - Em geral, os participantes entenderam o funcionamento da fase 1 da abordagem e conseguiram enxergar como a análise de *logs* pode trazer bons *insights* para a construção de questionários e entrevistas, e ainda servir de base para as recomendações de modernização do sistema. Durante a entrevista, mencionaram o fato de que, como os programas estavam escritos em python, seria necessário um recurso com conhecimento dessa tecnologia para realizar as execuções e novas análises que venham a ser necessárias. Desta forma, os programas de análise de *logs* deveriam ser mais genéricos e mais bem documentados, visando a facilitar o seu uso.

Participante 2 - “Achei a abordagem da fase bem interessante. Por meio dos *logs*, conseguimos ter uma boa ideia do comportamento das pessoas que acessam o sistema. Com esses dados surgem algumas ideias, e é possível elaborar um planejamento inicial para a próxima fase.”

Participante 3 - “Bem interessante, muitos *logs* onde conseguimos obter informações proveitosas para análises. Gráficos claros e objetivos que ajudam na análise”.

- **Fase 2** - Durante a entrevista, também não foram encontrados obstáculos para a execução da fase. Os participantes ficaram bastante surpresos de ver como os resultados se conectaram de formas diferentes: por um lado a

informação sobre a importância das notícias atuais que impactam o clube foi confirmada, e por outro lado, mesmo que o maior fluxo de acesso do site seja pelos links externos, o questionário indica que mais de 30% dos usuários que responderam nem sabe que os links existem.

Participante 3 - “O que achei mais interessante nesta fase é poder confrontar o resultado dos *logs* com os questionários e entrevistas. Isso pode enriquecer bastante o trabalho e contribuir para uma melhor avaliação da percepção dos usuários.”

Participante 3 - “Interessante o confronto entre os *logs* e o resultado do questionário, pois podemos verificar pontos de melhorias, por exemplo, nos temas referentes aos *links* e as pessoas que responderam péssimo”

- **Fase 3** - Os dois participantes se sentiram confortáveis em realizar as sugestões de modernização com os resultados encontrados nas duas primeiras fases.

Participante 2 - “Com tudo que foi demonstrado sobre a fase, me sinto confortável em recomendar este método para iniciar as recomendações sugeridas. As informações coletadas são extremamente úteis para iniciar as mudanças.”

Participante 3 - “Com o que foi indicado como resultado das fases 1 e 2, fica tranquilo analisar e sugerir recomendações de modernização.”

- **Fase 4** - Os participantes entenderam que a aplicação da fase 4 pode ocorrer sem contratempos, uma vez que é uma repetição das três primeiras fases com foco no sistema modernizado.
- **Dificuldades** - Ambos os participantes reportaram que não tiveram dificuldades de entendimento e utilização da abordagem.
- **Impressões sobre os códigos** - Não foi possível avaliar o código python. Porém, conforme dito na fase 1, seria necessária a participação de desenvolvedores com perfil nesta tecnologia.
- **Mais gostou - Participante 2** - “De que é uma abordagem que pode ser utilizada em diversos sistemas diferentes (independente de tecnologia ou área fim).” **Participante 3** - “A abordagem se mostra flexível para ser aproveitada em outros sistemas legados, além de ser de fácil entendimento”.
- **Menos gostou** - O fato de ter que encontrar um desenvolvedor da tecnologia pode ser um entrave para o uso.
- **Sugestões de Melhoria** - Não foram colocadas sugestões de melhoria neste momento.

- **Avaliação Final** - Os dois participantes consideraram que a abordagem tem grande potencial de ser utilizada em diferentes sistemas, entendem que a aplicação da abordagem não é um desafio e que o único ponto que precisa de atenção é a necessidade de um desenvolvedor com perfil específico de python para a manutenção dos programas de análise de *logs*.

6.5.3

Resumo

Os resultados demonstraram que a utilização da abordagem por terceiros é viável, uma vez que as fases estão bem definidas e não apresentaram nenhum problema de entendimento.

Um ponto de atenção se refere aos programas de análise de *logs*. Para facilitar a sua utilização, eles devem se tornar mais genéricos a fim de abranger outros tipos de sistema legado em outra linguagem de programação. Além disso, o formato dos *logs* deve ser mais bem documentado. Isso vai tornar a abordagem mais acessível a desenvolvedores com pouca experiência na tecnologia utilizada.

6.5.4

Respostas às Questões de Pesquisa

Retomamos aqui as questões de pesquisa desta dissertação e discutimos como a abordagem e sua aplicação contribui para respondê-las. Vale observar que todas as fases da abordagem contribuem de alguma forma para todas as perguntas.

RQ1: Como realizar uma modernização de sistemas legados a fim de melhorar a experiência do usuário sem perda de conhecimento de negócio com baixo custo e baixo esforço?

A abordagem possibilitou identificar que a análise de *logs*, questionários e entrevistas consegue prover informações suficientes para a realização de recomendações que melhorem a experiência do usuário sem mexer em regras de negócio, com baixo custo e baixo esforço de desenvolvimento. As sugestões formuladas no objeto de estudo foram realizadas na camada externa do sistema e trouxeram resultados importantes no engajamento e na percepção de um sistema mais interativo e moderno por parte do usuário.

RQ2: De que maneira a alteração de uma interface a partir de sugestões de modernização levantadas pela análise de *logs*, questionários e entrevistas contribui para influenciar a uma melhor experiência do usuário e seu engajamento dentro do sistema?

A abordagem possibilitou termos acesso eficiente aos dados de navegação dos usuários e trouxe recomendações que aumentaram o engajamento do usuário, trouxeram facilidades de acesso a informações que ou estavam escondidas ou traziam algum grau de dificuldade para a sua extração, com custo zero na utilização de ferramentas e com baixo esforço de um único recurso não dedicado ao desenvolvimento. Tudo isso ocorreu alterando-se somente a “casca” do sistema, mantendo o funcionamento das consultas (regras de negócio) sem alteração. A criação dos sistemas de recomendação sugeridos pela abordagem e desenvolvidos e implementados no sistema legado foi responsável por um aumento de 350% no engajamento da tela principal, além de um aumento de 20% no engajamento total do sistema, além de ter um reconhecimento de uma melhor experiência no acesso de 95% dos usuários que responderam o questionário.

7

Conclusões

Nesta dissertação, propusemos uma abordagem de modernização de sistemas legados com base em análise de *logs*, mantendo as funcionalidades principais do sistema sem alteração. Para avaliar a proposta, escolhemos como objeto de estudo um website, <https://www.fluzao.info>, e aplicamos a abordagem proposta, utilizando o Google Analytics como ferramenta para a geração dos *logs*. Criamos programas para análise dos *logs* gerados, aplicamos questionários e realizamos entrevistas com usuários. Com base nos dados coletados, sugerimos e implementamos melhorias e realizamos novas análises de *logs* específicas para as melhorias implementadas. Finalmente, avaliamos a abordagem com analistas externos.

Este capítulo apresenta como a abordagem atingiu os objetivos traçados nas questões de pesquisa, as principais contribuições e traça rumos para trabalhos futuros.

7.1

Questões de Pesquisa

RQ1: Como modernizar sistemas legados através de uma análise de *logs*, questionários e entrevistas?

Resultados - A abordagem trouxe resultados promissores na modernização do site <https://www.fluzao.info>. Tivemos um aumento de engajamento no acesso ao sistema, mais usuários acessaram o sistema depois das ações em redes sociais e modernizações nas telas de consulta e exibição. Houve também um aumento na exploração da página inicial com a implantação de um sistema de recomendação por conteúdo baseado nas últimas informações importantes relacionadas ao clube.

RQ2: Como a alteração de uma interface através de sugestões levantadas pela análise de *logs*, questionários e entrevistas pode contribuir para uma melhor experiência do usuário?

Resultados - Números mais altos de engajamento no sistema e na tela inicial, modernizada com a implantação de um sistema de recomendação. Mais usuários acessaram o sistema, mesmo em meio a um cenário desfavorável, com o clube passando por uma crise esportiva em 2024.

Obtivemos um elevado índice de aprovação das novas telas, resultado coletado pelo segundo questionário. Esses foram os principais indicativos do atingimento do objetivo traçado com a abordagem, mostrando que a experiência do usuário se tornou mais satisfatória.

As páginas e informações que não vinham sendo acessadas com regularidade passaram a ser acessadas, com origem do acesso no sistema de recomendação, nos links internos ou até mesmo nas consultas originais, reagrupadas e sinalizadas dentro do sistema, mostrando que a navegação se tornou mais eficiente.

7.2

Contribuições

As principais contribuições deste trabalho são a abordagem de modernização de sistemas legados e um sistema de análise de *logs* para esse tipo de sistema.

7.2.1

Abordagem de Modernização

A abordagem de modernização de sistemas legados através de análise de *logs*, questionários e entrevistas procura transformar o processo de modernização, normalmente custoso, complexo e caótico, em algo mais suave, simples e com menos custos, proporcionando uma nova experiência de UX/UI ao usuário, mantendo a confiabilidade do sistema.

Ao final do processo, o sistema legado ganha uma base de documentação que continuará sendo útil em manutenções futuras, o que torna esse aspecto da abordagem proposta uma contribuição importante.

A abordagem aplicada a um site de busca de informações apresentou resultados positivos. A etapa inicial de análise de *logs* trouxe *insights* de navegação que permitiram a elaboração das sugestões de forma clara. Em sua maioria, a adequação dessas sugestões foi confirmada com a aplicação dos questionários nos usuários do site. As recomendações sugeridas e implementadas foram avaliadas como exitosas, conforme evidenciado na continuação da análise de *logs* e no questionário e entrevista realizado com os usuários do sistema.

7.2.2

Sistema para Análise de *logs*

O sistema para análise de *logs* criado para análise de sistema legado pode ser reutilizado para outros modelos de modernização de sistema legado, podendo também ser utilizado para outros fins que não sejam modernização de sistema.

7.3

Trabalhos Futuros

Como trabalhos futuros, temos a generalização dos programas de análise de *logs*, permitindo assim que a abordagem possa ser aplicada a outros tipos de sistemas e sites legados. No momento, temos programas que se aplicam bem para sites de recuperação de informação. Para sites transacionais, como de venda de produtos, os programas precisariam de ajustes. No futuro, um processo de generalização poderá facilitar a utilização da abordagem.

Através de um trabalho futuro de generalização, será possível aplicar a abordagem a outros tipos de sistema legado, permitindo que sistemas importantes para as organizações mantenham as suas funcionalidades de forma mais moderna e com maior qualidade de uso para os usuários, com baixo custo e baixo esforço de evolução.

Para sistemas legados não baseados em sites, sistemas legados com telas próprias e navegação simples e repetitiva, os programas de análise de *logs* também devem ser adaptados para permitir identificar pontos relevantes e, conseqüentemente, fazer recomendações adequadas.

Referências Bibliográficas

- Abdellatif, M., Shatnawi, A., Mili, H., Moha, N., Boussaidi, G. E., Hecht, G., Privat, J., e Guéhéneuc, Y.-G. (2021). A taxonomy of service identification approaches for legacy software systems modernization. *Journal of Systems and Software*, 173:110868.
- Allam, S. (2018). An Exploratory Survey of Hadoop Log Analysis Tools.
- Althani, B. e Khaddaj, S. (2017). Systematic Review of Legacy System Migration. In *Proceedings - 2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science, DCABES 2017*, volume 2018-September, pages 154–157.
- Apaolaza, A. e Vigo, M. (2017). WevQuery: Testing Hypotheses about Web Interaction Patterns. *Proceedings of the ACM on Human-Computer Interaction*, 1(EICS):4:1–4:17.
- Baghdadi, Y. e Al-Bulushi, W. (2015). A guidance process to modernize legacy applications for SOA. *Service Oriented Computing and Applications*, 9(1):41–58.
- Bakar, H., Razali, R., e Jambari, D. (2019). Implementation phases in modernisation of legacy systems. In *International Conference on Research and Innovation in Information Systems, ICRIS*, volume December-2019.
- Bakar, H. K. A. e Razali, R. (2013). A preliminary review of legacy information systems evaluation models. In *2013 International Conference on Research and Innovation in Information Systems (ICRIIS)*, pages 314–318.
- Bianchi, A., Caivano, D., Marengo, V., e Visaggio, G. (2003). Iterative reengineering of legacy systems. *IEEE Transactions on Software Engineering*, 29(3):225–241.
- Bowen, J. (2015). Creating models of interactive systems with the support of lightweight reverse-engineering tools. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '15, pages 110–119, New York, NY, USA. Association for Computing Machinery.

- Breitmayer, M., Arnold, L., Rocca, S., e Reichert, M. (2023). Deriving Event Logs from Legacy Software Systems. *Lecture Notes in Business Information Processing*, 468 LNBIP:409–421.
- Brown, A. J. (1993). Specifications and reverse-engineering. *Journal of Software Maintenance: Research and Practice*, 5(3):147–153.
- Buschek, D., Eiband, M., e Hussmann, H. (2022). How to Support Users in Understanding Intelligent Systems? An Analysis and Conceptual Framework of User Questions Considering User Mindsets, Involvement, and Knowledge Outcomes. *ACM Transactions on Interactive Intelligent Systems*, 12(4):29:1–29:27.
- Chaimalas, I., Walker, D. M., Gruppi, E., Clark, B. R., e Toni, L. (2023). Bootstrapped Personalized Popularity for Cold Start Recommender Systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*, RecSys '23, pages 715–722, New York, NY, USA. Association for Computing Machinery.
- De Lucia, A., Francese, R., Scanniello, G., e Tortora, G. (2008). Developing legacy system migration methods and tools for technology transfer. *Software - Practice and Experience*, 38(13):1333–1364.
- De Vargas Agilar, E., De Almeida, R., e Canedo, E. (2016). A systematic mapping study on legacy system modernization. In *Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE*, volume 2016-January, pages 345–350.
- Dodero, J., Rodriguez-Garcia, M., Ruiz-Rube, I., e Palomo-Duarte, M. (2019). Privacy-preserving reengineering of model-view-controller application architectures using linked data. *Journal of Web Engineering*, 18(7):695–728.
- Feathers, M. (2004). *Working effectively with legacy code*. Pearson, 1 edition.
- García-Borgoñón, L., Barcelona, M., Egea, A., Reyes, G., Sainz-de-la-maza, A., e González-Uzabal, A. (2023). Lessons Learned in Model-Based Reverse Engineering of Large Legacy Systems. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13901 LNCS:330–344.
- Herbst, H., Knolmayer, G., Myrach, T., e Schlesinger, M. (1994). The specification of business rules: A comparison of selected methodologies. In *Methods and associated tools for the information systems life cycle*, pages 29–46. Citeseer.

- Hussain, S., Bhatti, S., e Rasool, M. (2017). Legacy system and ways of its evolution. In *International Conference on Communication Technologies, ComTech 2017*, pages 56–59.
- Jayathilake, D. (2012). Towards structured log analysis. In *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*, pages 259–264.
- Jomhari, N., Ahmad Alias, N. A., Ellah, A. A. A., Magableh, A. A., e Ghazali, E. M. (2024). A Multi-Criteria Decision-Making for Legacy System Modernization With FUCOM-WSM Approach. *IEEE Access*, 12:48608–48619.
- Kazanavičius, J. e Mažeika, D. (2023). An Approach to Migrate from Legacy Monolithic Application into Microservice Architecture. In *2023 IEEE Open Conference of Electrical, Electronic and Information Sciences, eStream 2023 - Proceedings*.
- Klammer, C. e Pichler, J. (2014). Towards tool support for analyzing legacy systems in technical domains. In *2014 Software Evolution Week - IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering (CSMR-WCRE)*, pages 371–374.
- Leon, P. e Horita, F. (2021). On the modernization of systems for supporting digital transformation: A research agenda. In *ACM International Conference Proceeding Series*.
- Lubos, S., Le, V.-M., Felfernig, A., e Tran, T. N. T. (2023). Analysis Operations for Constraint-based Recommender Systems. In *Proceedings of the 17th ACM Conference on Recommender Systems, RecSys '23*, pages 709–714, New York, NY, USA. Association for Computing Machinery.
- Mateos, C., Zunino, A., Flores, A., e Misra, S. (2019). COBOL systems migration to SOA: Assessing antipatterns and complexity. *Information Technology and Control*, 48(1):71–89.
- Medeiros, H., Batista, T., e Cavalcante, E. (2023). On a Metaprocess for Microservice Migration. In *ACM International Conference Proceeding Series*, pages 116–121.
- Miransky, A., Hamou-Lhadj, A., Cialini, E., e Larsson, A. (2016). Operational-Log Analysis for Big Data Systems: Challenges and Solutions. *IEEE Software*, 33(2):52–59.

- Noffsinger, W. B., Niedbalski, R., Blanks, M., e Emmart, N. (1998). Legacy object modeling speeds software integration. *Communications of the ACM*, 41(12):80–89.
- Pellegrini, R., Zhao, W., e Murray, I. (2022). Don't recommend the obvious: Estimate probability ratios. In *Proceedings of the 16th ACM Conference on Recommender Systems*, RecSys '22, pages 188–197, New York, NY, USA. Association for Computing Machinery.
- Pereira, C., Martinez, L., e Favre, L. (2023). Toward an Execution Trace Standard: An MDE Approach. *Communications in Computer and Information Science*, 1778 CCIS:91–106.
- Ponnusamy, S. e Eswararaj, D. (2023). Navigating the Modernization of Legacy Applications and Data: Effective Strategies and Best Practices. *Asian Journal of Research in Computer Science*, 16(4):239–256.
- Rashed, A., Elsayed, S., e Schmidt-Thieme, L. (2022). Context and Attribute-Aware Sequential Recommendation via Cross-Attention. In *Proceedings of the 16th ACM Conference on Recommender Systems*, RecSys '22, pages 71–80, New York, NY, USA. Association for Computing Machinery.
- Sneed, H. (2009). A pilot project for migrating COBOL code to web services. *International Journal on Software Tools for Technology Transfer*, 11(6):441–451.
- Sneed, H. e Verhoef, C. (2019). Re-implementing a legacy system. *Journal of Systems and Software*, 155:162–184.
- Sneed, H. M. (2000). [No title found]. *Annals of Software Engineering*, 9(1/4):293–313.
- Somogyi, N. e Kovesdan, G. (2021). Software Modernization Using Machine Learning Techniques. In *SAMI 2021 - IEEE 19th World Symposium on Applied Machine Intelligence and Informatics, Proceedings*, pages 361–365.
- Strobl, S., Bernhart, M., e Grechenig, T. (2020). Towards a Topology for Legacy System Migration. In *Proceedings - 2020 IEEE/ACM 42nd International Conference on Software Engineering Workshops, ICSEW 2020*, pages 586–594.
- Taveter, K. e Wagner, G. (2001). Agent-oriented enterprise modeling based on business rules. In *International Conference on Conceptual Modeling*, pages 527–540. Springer.

- Trabelsi, I., Abdellatif, M., Abubaker, A., Moha, N., Mosser, S., Ebrahimi-Kahou, S., e Guéhéneuc, Y.-G. (2023). From legacy to microservices: A type-based approach for microservices identification using machine learning and semantic analysis. *Journal of Software: Evolution and Process*, 35(10).
- Visaggio, G. (2001). Ageing of a data-intensive legacy system: Symptoms and remedies. *Journal of Software Maintenance and Evolution: Research and Practice*, 13(5):281–308.
- Zhu, Z. e Van Roy, B. (2023). Deep Exploration for Recommendation Systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*, RecSys '23, pages 963–970, New York, NY, USA. Association for Computing Machinery.

A

Material do Estudo

O material do estudo consistiu num Termo de Consentimento Livre e Esclarecido (TCLE – seção A.1), um questionário de caracterização do perfil do participante e captura de opiniões sobre o sistema (seção A.2).

A.1

Termo de Consentimento Livre e Esclarecido

Você está sendo convidado a participar da pesquisa que envolve o reprojeto do site "fluzao.xyz", sob a responsabilidade de Rodrigo Lima (fluzaoxyz@gmail.com). Este questionário tem o objetivo de compreender os interesses dos usuários do site e identificar oportunidades de melhoria.

O questionário é anônimo e as informações que você compartilhar conosco serão mantidas confidenciais e em pasta protegida por senha. Ninguém fora da equipe de pesquisa e desenvolvimento do site terá acesso às suas respostas ou mesmo saberá que você participou deste estudo.

Você deve ter ao menos 18 anos de idade para participar da pesquisa. Caso você tenha alguma dúvida ou preocupação com relação aos seus direitos como participante do estudo, você pode nos contactar por e-mail.

Sua participação é voluntária e você pode interromper o questionário a qualquer momento. Ao completar o questionário, você estará consentindo com o uso dos dados coletados para fins desta pesquisa.

A.2

Questionário para Captura de Opiniões sobre o Sistema

Perguntas - https://docs.google.com/forms/d/e/1FAIpQLSdY9xWWXCrtxeAB4rvM44aSu/viewform?usp=sf_link

Respostas - <https://docs.google.com/forms/d/1WckvhxNJwIwD0o10VFGGxiNsA0bY3zjT5pNzTI/edit#responses>

A.3

Arquivos de Logs coletados e analisados

https://github.com/RodLima26/Analise_legacy

A.4**Documento de Sugestão**

https://docs.google.com/document/d/1IVx1drjwZGGM8gwSfZdQuCIAzCENFt4I/edit?usp=drive_link&ouid=110656767057205552458&rtpof=true&sd=true

A.5**Documento de Sugestão Formulário**

https://docs.google.com/document/d/1IVx1drjwZGGM8gwSfZdQuCIAzCENFt4I/edit?usp=drive_link&ouid=110656767057205552458&rtpof=true&sd=true

A.6**Questionário para Captura de Opiniões sobre o Sistema Alterado**

Perguntas - https://docs.google.com/forms/d/e/1FAIpQLSfh7_Hzz1AfgkWOaZty1DL0maLxQX700UlpHoAGBxJiapretw/viewform?usp=sf_link

Respostas - <https://docs.google.com/forms/d/1fbUFi8N3JxwZVMdCMSLTXhKO-tdZ72-nk4NJpJXqcuo/edit#responses>

A.7**Texto da Abordagem para avaliação**

https://docs.google.com/document/d/1jf_V6STkd1XVsWDhjoUmhAa3cvRy820/edit?usp=drive_link&ouid=110656767057205552458&rtpof=true&sd=true

A.8**Roteiro de entrevista**

https://docs.google.com/forms/d/e/1FAIpQLSdj0C_6X2Q2jKDY0igBUPA_OfPaWekfsiUtV0Rwqy8zLBR5tg/viewform?usp=sf_link

A.9**Google colab usado nas avaliações de abordagem**

Análise Especiais - <https://colab.research.google.com/drive/16LcgD-eCRxu8FeJkEpA6r0GSNw-eMLQL> Análise Consultas Primárias - <https://colab.research.google.com/drive/1CWnbl28i2HTyu30sgL8pKyEwEj0svkHW>
Análise Consulta Zerada - https://colab.research.google.com/drive/12o_aISXFR0a2bN0It8NELx8sqlfjCDNO Análise Questionário - https://colab.research.google.com/drive/1h20vzzR2d_

wi1N8dst1ON3Px8LBL3Ipw Análise Tecnologia - <https://colab.research.google.com/drive/13r9N5IDt56wSYwKoEuRgPtkQRoaYG46c> Análise de Títulos - <https://colab.research.google.com/drive/1kdvjrnXyROfDtG7-08Z79YcDabW9ehoB> Análise de Páginas - <https://colab.research.google.com/drive/15SHWa2Een0jho00z741MwbNwKxdBPjC3> Análise de Atividades - https://colab.research.google.com/drive/10qj4sFso_3gMfFHA1mKXbHDkEL0Yk70Y Análise de Acessos - https://colab.research.google.com/drive/1kGp1PqodeNq_ss2h_CZ9xuWCwgklrP8V

B

Códigos Utilizados neste Trabalho

Este apêndice descreve os arquivos de código utilizados neste trabalho.

Os códigos referentes a abordagem de redesign de sistemas legados através da análise de logs estão no GitHub, em: https://github.com/RodLima26/Analise_legacy.

B.1

Informações sobre o código

Módulos

B.1.1

Gera referência

Gera_referencia.ipynb

Entradas – *referencia.csv* e *telas_acessadas.csv*

Saídas – *referencia_atualizada.csv*

Objetivo – Atualizar as colunas Visualizacao e Usuario com as quantidades de acesso, gerando o *referencia_atualizada.csv* que será a base de todo o resto da análise de páginas acessadas.

Detalhe específico do objeto de estudo - O campo código da *referencia.csv* é único no sistema legado e portanto é utilizado para atualizar os acessos de todas as páginas que contenham código no relatório gerado pelo analytics(*telas_acessadas.csv*), páginas de link e gráfico também são atualizadas com o código de acesso, as demais páginas são consideradas páginas de trânsito e não contém informações finais por isso não são consideradas importantes para o estudo e neste caso serão desconsideradas.

B.1.2

Análise de títulos por dia

Analise_Titulos.ipynb

Entradas – *Titulos_data.csv*

Saídas – Gráficos de análise de navegação, os gráficos que vão ser gerados dependem da análise de título que está sendo realizada

Objetivo – Realizar a análise da navegação dos usuários por data, fazendo o corte por data de acesso e por palavras-chave que agrupem consultas/páginas de mesma natureza para que seja possível estabelecer uma conexão dos acessos com eventos externos, permitindo a criação de um sistema de recomendação por conteúdo. Podem ser criados filtros por palavra-chave tantos quanto necessário para realizar a melhor análise. Outro objetivo é identificar as páginas mais acessadas do sistema legado independente de data e agrupamentos, outra análise executada é sobre os maiores acessos por data mas sem agrupamento e por último removendo algumas páginas, e.g. página inicial, para identificar algum viés de acesso.

Detalhe específico do objeto de estudo – Esta sendo analisado um sistema legado de estatística de um clube de futebol, então é importante analisar qual o adversário que tem mais acesso no geral, qual adversário que tem mais acesso perto dos confrontos, quais jogadores tem a sua página mais acessada, se existe alguma conexão com os dados informados em transmissões esportivas com as consultas realizadas em seguida.

B.1.3

Análise de títulos por hora

Analise_Titulos_Hora.ipynb

Entradas – *Titulos_data_hora.csv*

Saídas – Gráficos de análise de navegação, os gráficos que vão ser gerados dependem da análise de título que está sendo realizada

Objetivo – Realizar a análise da navegação dos usuários por data e hora, fazendo o corte por data de acesso e por palavras-chave que agrupem consultas/páginas de mesma natureza para que seja possível estabelecer uma conexão dos acessos com eventos externos, permitindo a criação de um sistema de recomendação por conteúdo. Podem ser criados filtros por palavra-chave tantos quanto necessário para realizar a melhor análise. Outro objetivo é identificar as páginas mais acessadas do sistema legado independente de data e agrupamentos, outra análise executada é sobre os maiores acessos por data mas sem agrupamento e por último removendo algumas páginas, e.g. página inicial, para identificar algum viés de acesso.

Detalhe específico do objeto de estudo – Este estudo foi criado com a intenção de identificar se durante os jogos o sistema recebia mais ou menos acessos, foi apurado que o sistema recebia menos acessos durante os jogos, porém seus picos são na noite anterior e logo após a disputa.

B.1.4

Análise de acesso a páginas por dia

Analise_Paginas.ipynb

Entradas – *referencia_atualizada.csv*; *jogo.csv*; *telas_acessadas.csv*

Saídas – Gráficos de análise de navegação, os gráficos que vão ser gerados dependem da análise de páginas, grupos/subgrupos e jogos que está sendo realizada

Objetivo – Realizar a análise da navegação nas páginas dos usuários por data, fazendo o corte por data de acesso, identificar páginas com pouco ou nenhum acesso, os grupos e subgrupos com maior quantidade relativa de consultas sem acesso, grupos e subgrupos com mais acessos totais.

Detalhe específico do objeto de estudo – Para o estudo em questão é importante identificar os jogos mais acessados, os títulos dos jogos nesse caso podem se repetir pelo fato de não ter a data como parte integrante, por isso usamos o código do jogo o que só é coletado na análise de páginas.

B.1.5

Análise dos acessos ao sistema

AnalisedeAcessos.ipynb

Entradas : *acessos.csv*; *acesso_cidades.csv*; *acesso_estado.csv*

Entradas Engajamento - *Engajamento.csv*; *Engajamento_tela1.csv*;

Saídas – Gráficos de quantidade de acessos e engajamento.

Objetivo – Análise dos acessos gerais, por cidade e por estado, buscando correlações entre os acessos e eventos externos, além do tempo de engajamento atual, um dos objetivos do estudo é aumentar este tempo com informações dispostas de forma mais inteligente pelo novo sistema. Análise de tempo de engajamento nos acessos a página principal do sistema utilizando o marco de implementação do sistema de recomendação.

Detalhe específico do objeto de estudo – Não existe nenhuma análise especial voltada diretamente para o objeto de estudo neste programa.

B.1.6

Análise de Títulos com pouco ou nenhum acesso

Analise_consulta_zerada.ipynb

Entradas – *referencia_atualizada.csv*; Saídas – *consultas_menos_acesso.csv*;

Gráficos de consultas zeradas ou com pouco acesso, grupos percentualmente com menos acesso.

Objetivo – Realizar a análise das consultas com pouco ou nenhum acesso, essa informação é base para a geração dos questionários e entrevistas.

Detalhe específico do objeto de estudo – Não temos nada específico neste programa.

B.1.7

Análise da tecnologia utilizada para acessar o sistema

Analise_tecnologia.ipynb

Entradas – *como_acessa_data.csv; onde_acessa_data.csv*

Saídas – Gráficos de Dispositivos e ferramentas de busca utilizadas para iniciar acesso ao sistema legado.

Objetivo – Realizar a análise dos dispositivos e ferramentas de busca, importante avaliar para identificar nível de responsividade exigido e necessidade de maior exposição em redes sociais e outros meios para aumentar o acesso e engajamento.

Detalhe específico do objeto de estudo – Não temos nada específico neste programa.

B.1.8

Identificação de links - Documentação

Identificando_links.ipynb

Entradas – biblioteca HTML do sistema legado

Saídas – Lista de links acessados pela tag *a* e lista de links acessados *viaurl_for*.

Objetivo – Gerar lista inicial de referência caso não exista uma documentação robusta com essa informação, identificar os links existentes no sistema para ajudar na recomendação de criação de novos links caso seja identificado que são insuficientes.

Detalhe específico do objeto de estudo – Não temos nada específico neste programa.

B.1.9

Análise de resposta aos questionários

Analise_questionario.ipynb

Entradas – Arquivo com o compilado das respostas do questionário.

Saídas – Gráficos gerados a partir da análise das respostas e correlações identificadas.

Objetivo – Gerar gráficos com as distribuições das respostas, analisar correlações entre respostas, identificar pontos comuns de melhoria pontados pelos usuários, gerar insumo para o roteiro de entrevista.

Detalhe específico do objeto de estudo – Não temos nada específico neste programa.

B.1.10

Análise de retorno das postagens em redes sociais

Analise_redes_sociais.ipynb

Entradas – Arquivo de títulos, arquivo de páginas e arquivo com resultado das redes sociais.

Saídas – Gráficos gerados a partir da análise das respostas e correlações identificadas.

Objetivo – Gerar gráficos com os resultados das postagens no X(Twitter), analisar os acessos as páginas citadas ou relacionadas nas postagens.

Detalhe específico do objeto de estudo – Não temos nada específico neste programa.

B.1.11

Análise de atividades realizadas no sistema

Analise_Atividades.ipynb

Entradas – Arquivo de atividades(page views, user engagements, scrolls, ...).

Saídas – Gráficos gerados a partir da análise das quantidades de atividades utilizando os marcos de atualização do sistema com base no documento de sugestões.

Objetivo – Gráficos gerados a partir da análise das quantidades de atividades utilizando os marcos de atualização do sistema com base no documento de sugestões.

Detalhe específico do objeto de estudo – Não temos nada específico neste programa.

B.1.12

Análise de acessos ao sistema de recomendação

Analise_Recomendacao.ipynb

Entradas – Arquivo de acessos página principal(Engajamento_tela1.csv).

Saídas – Gráficos gerados a partir da análise do tempo médio de engajamento por dia, por mês e percentual de aumento do tempo de engajamento.

Objetivo – Identificar a eficiência do sistema de recomendação com a maior retenção dos usuários no acesso ao sistema.

Detalhe específico do objeto de estudo – Não temos nada específico neste programa.

B.1.13

Análise de acessos as consultas primárias do sistema

Analise_Consultas_Primarios.ipynb

Entradas – Arquivo de acessos as consultas primárias principal(consulta_atributo.csv).

Saídas – Gráficos gerados a partir da análise número de usuários que acessaram, do tempo médio de engajamento por dia, por mês e percentual de aumento do tempo de engajamento.

Objetivo – Identificar a eficiência do ajuste realizado no procedimento de busca dos atributos primários do sistema.

Detalhe específico do objeto de estudo – Não temos nada específico neste programa.