



Trabalho de Conclusão de Curso

**Uma Abordagem Tecnológica para o Cumprimento  
da Lei Geral de Proteção de Dados: Para  
Especialistas e Leigos**

Lucas Buarque de Araujo Barros  
lbab@ic.ufal.br

Orientador:  
Prof. Dr. Ranilson Oscar Araújo Paiva

Maceió, Novembro de 2024

Lucas Buarque de Araujo Barros

# **Uma Abordagem Tecnológica para o Cumprimento da Lei Geral de Proteção de Dados: Para Especialistas e Leigos**

Monografia apresentada como requisito parcial para  
obtenção do grau de Bacharel em Engenharia de Com-  
putação pelo Instituto de Computação da Universidade  
Federal de Alagoas.

Orientador:

Prof. Dr. Ranilson Oscar Araújo Paiva

Maceió, Novembro de 2024

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**

Bibliotecária: Helena Cristina Pimentel do Vale – CRB4/661

B277u Barros, Lucas Buarque de Araujo.

Uma abordagem tecnológica para o cumprimento da lei geral de proteção de dados : para especialistas e leigos / Lucas Buarque de Araujo Barros. – 2024.  
59 f.: il.

Orientador: Ranilson Oscar Araújo Paiva.

Monografia (Trabalho de Conclusão de Curso em Engenharia da Computação) – Universidade Federal de Alagoas, Instituto de Computação. Graduação em Engenharia da Computação. Maceió, 2024.

Bibliografia: f. 57-59.

1. Lei geral de proteção de dados. 2. Sistemas de recomendação. 3. Web scraping.  
4. Web crawler. 5. Aplicação web. I. Título.

CDU: 004.77

# Agradecimentos

Quero expressar minha profunda gratidão aos meus pais por tudo o que fizeram por mim ao longo da minha trajetória. Agradeço por investirem nos meus estudos, acreditarem no meu potencial e estarem sempre ao meu lado em todos os momentos. Também sou muito grato à minha irmã Rebeca, que me incentivou a começar o curso de Engenharia de Computação e sempre trouxe sua criatividade para enriquecer meus projetos de graduação. Obrigado a todos vocês pelo apoio incondicional.

Gostaria de agradecer à minha namorada Bruna, que tem sido uma presença constante em minha vida desde o ensino médio. Sou profundamente grato pelo apoio que me ofereceu nos momentos mais difíceis, quando eu me sentia perdido, foi minha esperança e motivação para seguir em frente. Obrigado por estar sempre ao meu lado sendo minha companheira. Agradeço também à sua família pelo apoio incondicional que sempre me ofereceram.

Quero expressar minha gratidão aos irmãos que ganhei durante a graduação. Itallo Patrick, Jeferson Fernando e João Victor Holanda foram fundamentais para que eu pudesse enfrentar os desafios diários com mais leveza e alegria. O meu agradecimento especial ao Itallo que foi a minha dupla em todos os trabalhos em grupo e meu colega de trabalho nos projetos externos. Vocês são verdadeiros amigos para a vida toda.

Por fim, gostaria de agradecer ao professor Ranilson pelo apoio inestimável no desenvolvimento deste trabalho. Sou grato pela sua paciência e compreensão diante dos desafios que surgiram ao longo do processo. Sua orientação foi fundamental para que eu pudesse concluir esta jornada.

# Resumo

No cenário global, a segurança da informação é um tema de extrema relevância que está em constante debate. Com o avanço da tecnologia, grandes empresas coletam diariamente uma quantidade imensa de dados, levantando dilemas cruciais sobre a privacidade e os direitos dos titulares dessas informações. Esse aumento na coleta de dados não apenas facilita a personalização de serviços e produtos, mas também gera preocupações sobre como essas informações são armazenadas, utilizadas e compartilhadas. Assim, em 2018, o Brasil promulgou a Lei Geral de Proteção de Dados com o objetivo de assegurar os direitos dos indivíduos em relação aos seus dados pessoais, representando um marco importante na proteção da privacidade, garantindo maior transparência e controle aos titulares sobre suas informações. Dessa forma, os administradores de sites são legalmente obrigados a implementar as normas estabelecidas pela LGPD. Sem a orientação de um especialista, essa tarefa se torna cada vez mais desafiadora, exigindo conhecimentos técnicos e um entendimento aprofundado das exigências legais para garantir a conformidade e a proteção dos dados pessoais. Portanto, este trabalho tem como objetivo desenvolver uma ferramenta *web* de análise automática de *sites*, destinada a identificar não conformidades em relação à legislação e a sugerir ajustes necessários para assegurar a conformidade com a lei. Além disso, foi proposta uma comparação entre a análise realizada pela ferramenta e a análise manual realizada por voluntários que se dispuseram a participar de um experimento, oferecendo suas opiniões sobre a ferramenta.

**Palavras-chave:** Lei Geral de Proteção de Dados; Sistemas de recomendação; Web scraping; Web crawler; Aplicação web.

# Abstract

In the global landscape, information security is an extremely relevant topic that is constantly under debate. With the advancement of technology, large companies collect an immense amount of data daily, raising crucial dilemmas about privacy and the rights of data subjects. This increase in data collection not only facilitates the personalization of services and products but also generates concerns about how this information is stored, used, and shared. Thus, in 2018, Brazil enacted the General Data Protection Law (LGPD) with the aim of ensuring individuals' rights regarding their personal data, representing an important milestone in privacy protection by guaranteeing greater transparency and control for data subjects over their information. Consequently, website administrators are legally obligated to implement the standards established by the LGPD. Without the guidance of a specialist, this task becomes increasingly challenging, requiring technical knowledge and a deep understanding of legal requirements to ensure compliance and the protection of personal data. Therefore, this work aims to develop a web-based tool for automatic website analysis, intended to identify non-compliances with the legislation and suggest necessary adjustments to ensure compliance with the law. Additionally, a comparison was proposed between the analysis performed by the tool and the manual analysis conducted by volunteers who were willing to participate in an experiment, providing their feedback on the tool.

**Keywords:** General Data Protection Law; Recommendation systems; Web scraping; Web crawler; Web application.

# Lista de Figuras

2.1	Fases do processo de coleta de dados web (Fonte: Persson). . . . .	9
2.2	Exemplo da utilização da biblioteca BeautifulSoup (Fonte: Mitchell). . . . .	10
2.3	Diagrama da arquitetura limpa (Fonte: Martin). . . . .	11
3.1	Arquitetura para o desenvolvimento da aplicação (Fonte: Elaborada pelo autor).	15
3.2	Página inicial da aplicação (Fonte: Elaborada pelo autor). . . . .	17
3.3	Página sobre os critérios abordados (Fonte: Elaborada pelo autor). . . . .	18
3.4	Página de início do processamento (Fonte: Elaborada pelo autor). . . . .	19
3.5	Página de resultados do processamento (Fonte: Elaborada pelo autor). . . . .	20
3.6	Página de resultados do processamento (Fonte: Elaborada pelo autor). . . . .	21
3.7	Entidade de análise de um site (Fonte: Elaborada pelo autor). . . . .	23
3.8	Exemplo do mecanismo de expiração de dados (Fonte: red). . . . .	24
3.9	Imagem da interface do RabbitMQ (Fonte: Elaborada pelo autor). . . . .	24
3.10	Método que executa as tarefas do <i>crawler</i> (Fonte: Elaborada pelo autor). . . . .	26
3.11	Exemplo de uma tarefa de um <i>crawler</i> (Fonte: Elaborada pelo autor). . . . .	27
3.12	Instâncias dos <i>crawlers</i> (Fonte: Elaborada pelo autor). . . . .	27
3.13	Execução dos <i>crawlers</i> (Fonte: Elaborada pelo autor). . . . .	28
3.14	Implementação do consumidor (Fonte: Elaborada pelo autor). . . . .	28
5.1	Gráfico obtido a partir das respostas da questão sobre participação em pesquisas anteriores relacionadas à LGPD (Fonte: elaborada pelo autor). . . . .	39
5.2	Gráfico das respostas sobre experiência jurídica (Fonte: elaborada pelo autor). .	39
5.3	Gráfico das respostas sobre a familiaridade com a LGPD (Fonte: elaborada pelo autor). . . . .	40
5.4	Gráfico das respostas sobre a confiabilidade da análise manual (Fonte: elaborada pelo autor). . . . .	41
5.5	Gráfico das respostas sobre o auxílio da ferramenta na identificação de não conformidades com a LGPD. (Fonte: elaborada pelo autor). . . . .	43
5.6	Gráfico das respostas sobre o auxílio da ferramenta na melhoria da conformidade de um site (Fonte: elaborada pelo autor). . . . .	43
5.7	Gráfico das respostas sobre o auxílio da ferramenta na compreensão da LGPD (Fonte: elaborada pelo autor). . . . .	44
5.8	Gráfico das respostas sobre o benefício da ferramenta na manutenção da conformidade de um site (Fonte: elaborada pelo autor). . . . .	45
5.9	Gráfico das respostas sobre o auxílio da ferramenta em evitar multas por não conformidades em um site. (Fonte: elaborada pelo autor). . . . .	45
5.10	Gráfico das respostas sobre o auxílio da ferramenta no trabalho de um profissional de direito (Fonte: elaborada pelo autor). . . . .	46
5.11	Gráfico das respostas sobre a adesão do uso da ferramenta no trabalho de um profissional de direito (Fonte: elaborada pelo autor). . . . .	47

---

5.12	Gráfico das respostas sobre a substituição do especialista em direito pela ferramenta (Fonte: elaborada pelo autor). . . . .	47
5.13	Gráfico das respostas sobre como a ferramenta não substitui o especialista, mas complementa seu trabalho (Fonte: elaborada pelo autor). . . . .	48
5.14	Gráfico das respostas sobre a facilidade em aprender a usar a ferramenta (Fonte: elaborada pelo autor). . . . .	49
5.15	Gráfico das respostas sobre a facilidade de navegação na interface da ferramenta (Fonte: elaborada pelo autor). . . . .	50
5.16	Gráfico das respostas sobre o <i>feedback</i> imediato da ferramenta (Fonte: elaborada pelo autor). . . . .	50
5.17	Gráfico das respostas sobre a necessidade de treinamento para o uso da ferramenta (Fonte: elaborada pelo autor). . . . .	51
5.18	Gráfico das respostas sobre a acessibilidade da ferramenta (Fonte: elaborada pelo autor). . . . .	52
5.19	Gráfico das respostas sobre a necessidade de ser um profissional de direito para utilizar a ferramenta (Fonte: elaborada pelo autor). . . . .	52
5.20	Gráfico das respostas sobre a necessidade do conhecimento em programação para utilizar a ferramenta (Fonte: elaborada pelo autor). . . . .	53

# Conteúdo

<b>Lista de Figuras</b>	<b>iii</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Fundamentação Teórica</b>	<b>3</b>
2.1 Lei Geral de Proteção de Dados (LGPD)	3
2.1.1 Tipos de Dados Pessoais	4
2.1.2 Direitos do Titular dos Dados	4
2.2 Sistemas de Recomendação	5
2.2.1 Abordagens em Sistemas de Recomendação	5
2.2.2 Privacidade em Sistemas de Recomendação	6
2.3 <i>HyperText Transfer Protocol (HTTP)</i>	6
2.3.1 Cabeçalhos ( <i>Headers</i> )	6
2.3.2 Métodos	7
2.3.3 <i>Hypertext Markup Language (HTML)</i>	7
2.3.4 <i>Cookies</i>	8
2.4 Coleta de Dados Web ( <i>Web Scraping</i> )	8
2.4.1 Etapas do Processo	9
2.4.2 Coleta de Dados Web na Prática	10
2.5 Arquitetura Limpa ( <i>Clean Architecture</i> )	11
2.5.1 Entidades	12
2.5.2 Casos de Uso	12
2.5.3 Adaptadores de Interface	12
2.5.4 <i>Drivers e Frameworks</i>	13
<b>3 Proposta</b>	<b>14</b>
3.1 Arquitetura Geral do Projeto	15
3.2 Front-End	16
3.2.1 Tecnologias utilizadas	16
3.2.2 Telas do sistema	16
3.3 Back-end	21
3.3.1 Tecnologias utilizadas	22
3.3.2 Arquitetura de <i>software</i> utilizada	22
3.3.3 Entidades do sistema	22
3.3.4 Cache	23
3.4 Serviço de mensageria	24
3.5 Web Crawlers	25
3.5.1 Critérios avaliados	25

3.5.2	Desenvolvimento	26
3.6	Ambiente de produção	29
3.7	Objetivo geral	29
3.8	Objetivos específicos	29
<b>4</b>	<b>Metodologia</b>	<b>31</b>
4.1	Etapas da pesquisa	31
4.2	Seleção do público-alvo	32
4.3	Formulário de análise manual	32
4.3.1	Termo de Consentimento Livre e Esclarecido	32
4.3.2	Informações do participante	33
4.3.3	Análise manual	33
4.3.4	Fim do experimento	34
4.4	Definição do experimento	34
4.5	Pesquisa de satisfação	35
4.5.1	Percepção de utilidade	35
4.5.2	Facilidade de uso	36
4.6	Métricas de interesse	37
<b>5</b>	<b>Resultados e Discussões</b>	<b>38</b>
5.1	Perfil dos participantes	38
5.1.1	Participação em Pesquisas Relacionadas à LGPD	38
5.1.2	Experiência ou Trabalho na Área Jurídica	39
5.1.3	Nível de Familiaridade com a LGPD	39
5.1.4	Julgamento do Nível de Confiabilidade na Análise Manual	40
5.2	Tempo de análise	41
5.3	Precisão da Análise	41
5.4	Pesquisa de satisfação	42
5.4.1	Percepção de utilidade da ferramenta	42
5.4.2	Facilidade de uso da ferramenta	48
5.5	Discussões sobre os resultados	53
<b>6</b>	<b>Conclusão</b>	<b>55</b>
6.1	Trabalhos futuros	56
	<b>Referências bibliográficas</b>	<b>57</b>

# 1

## Introdução

Dado pode ser definido como uma sequência de símbolos que representam informações quantitativas ou qualitativas. Em outras palavras, dados são elementos brutos como números, textos, fotos, sons, que podem ser coletados, analisados e interpretados para gerar informação que, por sua vez, pode ser usada para apoiar ou tomar decisões [28]. Atualmente, os dados são considerados valiosos. Empresas, governos e instituições de diversos setores, como comercial, marketing, bancário e saúde, estão constantemente coletando consideráveis volumes de dados, para diferentes propósitos [6]. Nesse contexto, cada vez mais a vida cotidiana está sendo influenciada por algoritmos e programas de computador.

Assim, surgem preocupações com o uso ético, consciente e responsável desses dados, o que gera demanda para a criação de normas e regras que limitem esse uso. O principal objetivo das leis de proteção de dados é regular o uso e o tratamento de dados pessoais. Essas leis não só visam proteger a privacidade das pessoas, mas também asseguram outros direitos fundamentais e liberdades pessoais. Para garantir a plena execução desses direitos, é essencial assegurar o tratamento adequado dos dados pessoais, que frequentemente representam a identidade e as características dos indivíduos [21].

Diante disso, foi criada a Lei Geral de Proteção de Dados (LGPD, Lei 13.709/2018) com o objetivo de criar uma estrutura legal voltada especificamente para a proteção de dados pessoais [15]. Dessa forma, a adequação das empresas à LGPD torna-se obrigatória e está intimamente relacionada ao tema da segurança da informação, que abrange a proteção tanto dos dados da organização quanto dos seus clientes. Dentre os benefícios proporcionados pela lei, temos o fato de que as empresas não podem mais tratar os dados de terceiros de forma arbitrária, como era comum anteriormente [3].

Além disso, a implementação das normas estabelecidas tornou-se cada vez mais complexa e desafiadora [32], dificultando para as empresas que possuem aplicações *web* garantir a conformidade com a lei. Isso requer uma análise cuidadosa por parte de especialistas legais para assegurar que os *sites* estejam em conformidade com a legislação.

Tendo em vista a problemática abordada, o objetivo deste trabalho é desenvolver uma ferramenta para análise automática de não conformidades com a LGPD em páginas da *web*, com o intuito de auxiliar tanto administradores de *sites* quanto profissionais do direito na identificação dessas não conformidades. Além disso, o trabalho também visa obter a opinião de voluntários a respeito da ferramenta.

Dessa forma, os objetivos para este trabalho são:

- Propor uma arquitetura de *software* e escolher as tecnologias necessárias para o desenvolvimento de uma ferramenta (como prova de conceito) para a detecção automática, em *websites*, de não conformidades com a LGPD;
- Implementar a referida ferramenta;
- Realizar experimento utilizando a ferramenta, com a participação de voluntários (administradores de *websites* e especialistas na LGPD);
- Análisar os dados coletados durante o experimento e elaborar os resultados;
- Comparar os resultados obtidos com os resultados esperados.

Os capítulos posteriores estão dispostos da seguinte maneira:

- Capítulo 2: apresentará a fundamentação teórica que sustenta este trabalho;
- Capítulo 3: descreverá detalhadamente a proposta do trabalho;
- Capítulo 4: discutirá sobre as estratégias metodológicas adotadas;
- Capítulo 5: apresentará os resultados obtidos através do experimento com os voluntários;
- Capítulo 6: apresentará as conclusões do trabalho e as sugestões para futuros estudos.

# 2

## Fundamentação Teórica

### 2.1 Lei Geral de Proteção de Dados (LGPD)

Com o avanço e expansão global do acesso à tecnologia, crimes e violações de dados se tornaram mais comuns. No entanto, não havia uma legislação específica para proteger os cidadãos contra esses tipos de crime [26].

Na Europa, escândalos envolvendo espionagem e divulgação de dados de clientes, notadamente os casos envolvendo a *Cambridge Analytica* e o Facebook <sup>1</sup>, desencadearam uma ampla discussão sobre a proteção de dados pessoais. Como resultado desse debate, foi promulgado o Regulamento Geral de Proteção de Dados (GDPR), que estabelece normas abrangentes de segurança de dados dentro da União Europeia (UE). Essa legislação tem como objetivo garantir a privacidade dos cidadãos europeus e estabelecer padrões rigorosos para o tratamento de informações pessoais por empresas e organizações que operam na região [9].

Inspirada na lei europeia, em 14 de agosto de 2018, no Brasil, foi promulgada a Lei nº 13.709, conhecida como Lei Geral de Proteção de Dados Pessoais (LGPD), ou ainda, Marco Civil da Internet. Essa legislação foi concebida com o objetivo de estabelecer medidas preventivas e proativas para garantir a privacidade e segurança dos dados pessoais de terceiros [25].

Portanto, proprietários de *websites* em todo o país enfrentam a necessidade de ajustar suas páginas às recentes diretrizes legais, uma tarefa que tem se tornado cada vez mais desafiadora e abstrata. Nesse contexto, a ferramenta desenvolvida nesta pesquisa tem como finalidade automatizar a análise que seria feita através de uma avaliação minuciosa e manual por parte de um especialista na lei.

---

<sup>1</sup><https://olhardigital.com.br/2018/03/21/noticias/cambridge-analytica/>

### 2.1.1 Tipos de Dados Pessoais

Segundo a [Escola Superior do Ministério Público da União](#), a Lei Geral de Proteção de Dados classifica como dado pessoal qualquer informação que possa identificar uma pessoa natural. Esse tipo de dado pode ser dividido em algumas categorias:

- **Dados sensíveis:** qualquer informação de natureza pessoal relacionada à origem racial ou étnica, crença religiosa, opinião política, filiação a sindicatos ou organizações de caráter religioso, filosófico ou político, informações relacionadas à saúde ou à vida sexual, dados genéticos ou biométricos, quando associados a uma pessoa natural.
- **Dados anonimizados:** Relativos a titular que não possa ser identificado ou tornar-se identificável, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento.
- **Dados pseudonimizados:** quando um dado perde sua capacidade de ser associado, direta ou indiretamente, a um indivíduo, exceto quando são utilizadas informações adicionais mantidas separadamente pelo controlador em um ambiente seguro e controlado.
- **Outros dados pessoais:** dados usados para formar o perfil comportamental de uma pessoa natural, são considerados dados pessoais, especialmente quando essa pessoa pode ser identificada.

### 2.1.2 Direitos do Titular dos Dados

Define-se como titular de dados a pessoa física a quem os dados pessoais se referem e que está sujeita ao tratamento desses dados. Em outras palavras, é a pessoa natural sobre a qual os dados são coletados e processados [5]. Conforme [Rosal Santos](#), o titular dos dados detém direitos essenciais que fundamentam-se em três princípios-chave: transparência no manejo dos dados, garantia da segurança da informação e controle exercido pelo indivíduo sobre seus próprios dados pessoais.

O princípio da transparência, serve como base para diversos direitos listados na lei. Estes incluem a confirmação da execução do tratamento dos dados, a divulgação das entidades públicas e privadas com as quais o controlador compartilhou os dados e a informação facilitada sobre a possibilidade de não consentir e as consequências dessa recusa. Tais prerrogativas visam assegurar que o titular dos dados tenha acesso a informações que lhe permitam compreender as práticas da instituição.

A segurança da informação refere-se ao direito que o titular possui para solicitar a anonimização, bloqueio ou exclusão de dados que sejam excessivos ou tratados de forma inadequada conforme estipulado pela lei. Essa prerrogativa destaca a importância do equilíbrio no sistema de proteção de dados pessoais e garante que os dados serão utilizados para finalidades legítimas.

Em relação ao controle individual sobre os dados pessoais, [Rosal Santos](#) destaca que a legislação garante ao titular o direito de acessar as informações mantidas pela instituição, permitindo a correção de dados incompletos ou desatualizados e até mesmo a exclusão completa dos dados fornecidos com consentimento. Ademais, é facultado ao titular solicitar uma cópia de seus dados pessoais para transferi-los a outros fornecedores de serviço.

## 2.2 Sistemas de Recomendação

Segundo [Cazella et al.](#), a acessibilidade crescente a vastas quantidades de informação tem proporcionado às pessoas um aumento significativo no leque de opções disponíveis. No entanto, muitas vezes, os usuários enfrentam dificuldades devido à falta de experiência em relação às alternativas disponíveis. Assim, os sistemas de recomendação surgiram com o propósito de facilitar e otimizar o processo de sugestão de itens ao usuário, ampliando sua capacidade de fazer escolhas informadas e aumentando a eficácia das recomendações oferecidas.

### 2.2.1 Abordagens em Sistemas de Recomendação

De acordo com [Shah et al.](#), um sistema de recomendação pode ser implementado de duas maneiras distintas. Na abordagem inicial, o sistema oferece recomendações idênticas de itens avaliados para todos os usuários, sem levar em conta o perfil individual de cada usuário. Nesse tipo de sistema, as preferências individuais dos usuários não são consideradas, o que pode resultar em insatisfação.

Na outra abordagem, o sistema analisa as preferências individuais de cada usuário, levando em consideração diversos fatores como histórico de interações, avaliações prévias de itens e informações demográficas. Com base nesses dados, o sistema emprega algoritmos de recomendação que buscam identificar padrões e afinidades entre usuários e itens, com o objetivo de oferecer sugestões mais relevantes e personalizadas. Essa abordagem visa proporcionar uma experiência de recomendação mais eficaz, adaptada às necessidades e interesses específicos de cada usuário ou grupo de usuários.

No âmbito deste trabalho, a abordagem que não considera as preferências individuais foi escolhida. Esta escolha se deve ao fato de que o sistema realizará uma análise detalhada dos elementos presentes em uma página *web* específica, com base em critérios determinísticos, a fim de fornecer uma resposta precisa e previsível, automatizando a análise manual por parte de um profissional.

Visando aprimorar o desempenho da ferramenta em trabalhos futuros, é essencial incorporar um modelo de recomendação fundamentado na filtragem colaborativa. Dessa maneira, os usuários teriam a oportunidade de fornecer *feedback* sobre os resultados obtidos, contribuindo para ajustar as análises futuras e mitigar possíveis falhas.

## 2.2.2 Privacidade em Sistemas de Recomendação

A pesquisa personalizada na web representa uma abordagem promissora para aprimorar a qualidade das buscas, adaptando os resultados de acordo com os objetivos de informação individuais dos usuários. No entanto, muitos usuários expressam desconforto ao compartilhar informações de preferências pessoais com os mecanismos de busca, preocupados com a privacidade de seus dados. [33]

De acordo com [Cazella et al.](#) Dados comportamentais, como histórico de navegação e padrões de compra, juntamente com informações pessoais, são frequentemente coletados e armazenados pelos sistemas. Essa coleta, muitas vezes ocorre de maneira implícita, ou seja, sem que o usuário esteja necessariamente ciente de que suas informações estão sendo registradas enquanto ele utiliza o serviço. Portanto, é necessário que atrelado à coleta de dados exista uma política de privacidade objetiva e adequada.

Entretanto, [Cazella et al.](#) destaca que ainda é recorrente a prática de empresas coletarem dados de seus clientes e os comercializarem para terceiros, o que pode contribuir para a proliferação de práticas indesejadas, como o *spam*. Como resultado, os usuários dos sistemas de recomendação contemporâneos estão perdendo gradualmente a confiança nas empresas que coletam seus dados pessoais. Essa falta de confiança pode impactar negativamente a adoção e o engajamento dos usuários com esses sistemas, criando desafios adicionais para as empresas na gestão da privacidade e segurança dos dados.

## 2.3 *HyperText Transfer Protocol (HTTP)*

*HTTP* é um protocolo de aplicação para transferência de recursos, como documentos *HTML*, entre um cliente e um servidor web específico. Caracterizado por ser um modelo de comunicação sem estado, o protocolo atua seguindo o fluxo cliente-servidor. A comunicação entre esses clientes e os respectivos servidores web se dá a partir de solicitações e respostas *HTTP* [12].

Quando se trata de *websites*, o protocolo *HTTP* é o principal meio de comunicação. A ferramenta desenvolvida neste trabalho utiliza este protocolo através de requisições para obter os documentos *HTML*, os quais serão posteriormente analisados durante a fase de processamento.

### 2.3.1 Cabeçalhos (*Headers*)

De acordo com [Chapagain](#), dentro da implementação do protocolo existe o conceito de cabeçalhos. Eles possuem o objetivo de transferir informações adicionais para um determinado cliente ou servidor durante uma requisição ou resposta obtida. O seu conteúdo geralmente é

constituído por pares chave-valor de informações, as quais serão transferidas durante a comunicação por meio de dois tipos de cabeçalhos: solicitação e resposta.

- Solicitação: esse tipo de cabeçalho é utilizado durante as requisições do protocolo. Nele, poderá conter informações a respeito de idioma, codificação de caracteres, cookies, referenciadores e informações relacionadas ao navegador. Todos esses dados podem ser transmitidas ao servidor.
- Resposta: esse tipo de cabeçalho contém informações detalhadas acerca da resposta proveniente do servidor web. Em geral, ele incorpora dados relativos ao estado do servidor, tamanho do conteúdo da resposta e a data em que a mesma foi gerada. Essas informações desempenham um papel fundamental na compreensão e interpretação das respostas fornecidas pelos servidores web.

### 2.3.2 Métodos

Segundo [Gourley and Totty](#), os métodos são convenções criadas para padronizar determinadas ações que o cliente deverá informar para que o servidor as execute. Dentre os diversos métodos existentes no protocolo, os mais utilizados são: GET, HEAD, POST, PUT, DELETE, TRACE, OPTIONS.

- GET: é utilizado para obter um documento do servidor.
- HEAD: é utilizado apenas para obter os cabeçalhos de um documento do servidor.
- POST: método comumente usado para o envio de dados ao servidor visando o processamento dos mesmos.
- PUT: possui o objetivo de armazenar o corpo da requisição no servidor.
- DELETE: visa excluir algum documento armazenado no servidor.
- TRACE: é utilizado para rastrear a mensagem através de servidores de proxy até o servidor final.
- OPTIONS: tem como objetivo informar quais métodos podem ser utilizados no servidor.

### 2.3.3 *Hypertext Markup Language (HTML)*

A *Hypertext Markup Language (HTML)* é uma linguagem amplamente empregada na web para a marcação de hipertexto. Hipertexto refere-se ao conteúdo presente em um documento que se destaca por sua característica fundamental de interconectar-se com outros documentos disponíveis [30].

Além dessa característica, a linguagem possui o conceito de universalidade através dos documentos que são salvos em simples arquivos de texto. Com isso, é possível que qualquer computador leia e processe páginas na *web* independente de seu sistema operacional ou arquitetura [10].

É importante destacar que o protocolo *HTTP* não se limita à transferência exclusiva de documentos na linguagem de marcação *HTML*. Embora a compreensão da linguagem seja requerida por parte de um cliente do protocolo, o mesmo é utilizado na recuperação de documentos em uma ampla variedade de formatos, ilimitados e expansíveis [4].

### 2.3.4 *Cookies*

Os *cookies* são mecanismos de navegação *web* que foram criados com o objetivo de possibilitar o gerenciamento de estados entre clientes e servidores. Consistem em simples textos armazenados no navegador do cliente e no servidor específico. Essas informações são transmitidas para o servidor por meio dos cabeçalhos, onde serão posteriormente processadas [8].

Embora a concepção do uso dessas ferramentas seja vantajosa tanto para os usuários quanto para os desenvolvedores, a prática de técnicas manipuladoras, com o propósito de impactar os usuários das atuais aplicações, tem se tornado cada vez mais comum no cenário digital contemporâneo [24].

Portanto, a utilização dos *cookies* colabora com empresas que visam a distribuição de anúncios em diversos *sites* com o objetivo de implantar e posteriormente recuperar as informações dos dispositivos dos clientes. Dessa forma, ao analisar os dados obtidos do usuário, torna-se viável identificar as suas preferências, monitorando suas atividades, coletando informações pessoais e supervisionando compras ou contratações de serviços. Observando os mecanismos criados para o monitoramento através dessa ferramenta, é possível perceber comprometimento e riscos à privacidade do usuário [22].

## 2.4 Coleta de Dados Web (*Web Scraping*)

A coleta de dados tornou-se uma ferramenta de extrema importância na sociedade moderna. A obtenção e a análise de dados são elementos essenciais em pesquisas, sejam elas de caráter científico, de marketing ou acadêmico. As pessoas estão cada vez mais interessadas em coletar e analisar dados provenientes de páginas através da *web* [31].

No entanto, os variados *websites* presentes na Internet disponibilizam informações de diversas maneiras, e nem sempre é viável extrair todos os tipos de dados apenas por meio da navegação convencional. Diante disso, o método de *Web Scraping* surgiu com o propósito de extrair informações de múltiplos *sites* e consolidá-las em um único banco de dados ou planilha. Isso simplifica significativamente o processo de análise dos dados coletados [31].

Este método de extração desempenha um papel fundamental nesta pesquisa. Ele permite a obtenção de uma variedade de informações relevantes contidas em documentos *HTML*, explorando a estrutura hierárquica dos elementos por meio do Modelo de Objeto de Documentos (*DOM*).

### 2.4.1 Etapas do Processo

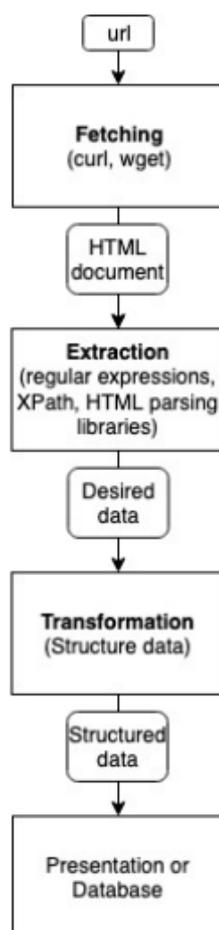


Figura 2.1: Fases do processo de coleta de dados web (Fonte: [Persson](#)).

De acordo com [Persson](#), o processo coleta dos dados de um *website* se divide em três etapas: busca, extração e transformação.

- Busca: Na etapa de busca, é necessário que o executor acesse o site que deseja extrair o seu conteúdo. Este acesso é feito através do protocolo *HTTP* (*Hypertext Transfer Protocol*), o qual é um protocolo da *internet* fundamentado no fluxo cliente-servidor, utilizando requisições e respostas através de um cliente e um servidor web. Essa abordagem é semelhante àquela realizada pelos navegadores *web* quando se comunicam com *sites*, proporcionando aos usuários uma experiência visual agradável. No entanto, é possível

realizar esse acesso por meio de outras ferramentas de comunicação entre protocolos de rede, como Netcat<sup>2</sup>, curl<sup>3</sup> e wget<sup>4</sup>. Dessa forma, essas ferramentas emitem requisições ao servidor web de maneira análoga aos navegadores, porém, a resposta consiste no próprio código *HTML*, sem o processamento realizado pelo navegador.

- **Extração:** Na fase de extração, o executor já dispõe do documento *HTML* coletado na etapa anterior. Contudo, para obter os dados desejados, é necessário realizar a extração propriamente dita. Através de técnicas como o uso de expressões regulares e bibliotecas de análise e interpretação sintática do código *HTML*, é possível programar uma série de regras que permitam ao realizador extrair as informações de interesse de forma eficiente.
- **Transformação:** Na etapa de transformação, o executor, tendo extraído com sucesso os dados de interesse, necessita apenas estruturá-los de maneira clara e objetiva, visando armazená-los em um banco de dados ou apresentá-los em tabelas para facilitar a visualização e gerar conhecimento.

## 2.4.2 Coleta de Dados Web na Prática

A linguagem Python<sup>5</sup> destaca-se como uma das ferramentas mais empregadas no desenvolvimento de *scripts* para extração e coleta de dados de *websites*. Sua popularidade é evidente através de uma comunidade robusta e ativa, dedicada ao ensino e à resolução de desafios enfrentados pelos programadores. Além disso, ela se destaca pela sua facilidade de leitura e compreensão devido à sintaxe simplificada e à abstração de operações complexas [18].

No âmbito de sua extensa variedade de bibliotecas, a BeautifulSoup<sup>6</sup>, focada na minuciosa etapa de extração, conforme delineado por Persson. De acordo com o descrito por Mitchell, esta realiza a tentativa de conferir significado ao que inicialmente pode parecer desordenado, contribuindo para a formatação e organização dos dados extraídos do documento *HTML* em objetos facilmente manipuláveis no código do *script*.

```
from urllib.request import urlopen
from bs4 import BeautifulSoup

html = urlopen('http://www.pythonscraping.com/pages/page1.html')
bs = BeautifulSoup(html.read(), 'html.parser')
print(bs.h1)
```

Figura 2.2: Exemplo da utilização da biblioteca BeautifulSoup (Fonte: Mitchell).

<sup>2</sup><https://netcat.sourceforge.net/>

<sup>3</sup><https://curl.se/>

<sup>4</sup><https://www.gnu.org/software/wget/>

<sup>5</sup><https://www.python.org/>

<sup>6</sup><https://www.crummy.com/software/BeautifulSoup/>

Na Figura 2.2, é possível visualizar a facilidade de entendimento e utilização da biblioteca em apenas poucas linhas de código. Nesse pequeno trecho, já é possível extrair todas as tags *HTML H1* existentes no conteúdo do *website* analisado.

## 2.5 Arquitetura Limpa (*Clean Architecture*)

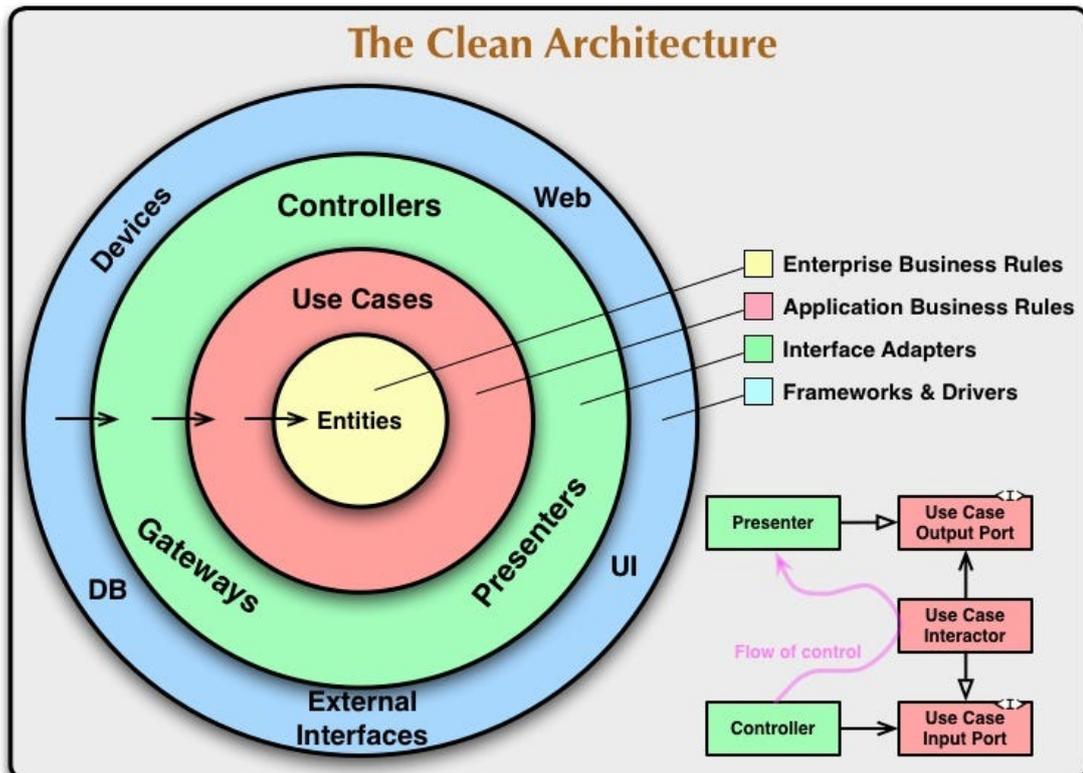


Figura 2.3: Diagrama da arquitetura limpa (Fonte: [Martin](#)).

A arquitetura limpa é um padrão arquitetural de desenvolvimento de *software*. Proposta por Robert Martin, mais conhecido como *Uncle Bob*, em seu livro *Clean Architecture: A Craftsman Guide to Software Structure and Design*, possui o objetivo de auxiliar a criação de sistemas independentes de qualquer tecnologia, facilmente testáveis, com camadas internas altamente desacopladas.

Segundo [Martin](#), as regiões mais externas do diagrama apresentado na Figura 2.3 são mecanismos e as mais internas são políticas. Nada que se encontra em uma camada mais interna poderá ter conhecimento de algum trecho de código nas camadas externas, incluindo funções, classes, variáveis ou qualquer outra entidade de *software* nomeada. Sendo assim, componentes internos devem depender de interfaces bem definidas, em vez de detalhes sobre a implementação.

Para o desenvolvimento da ferramenta neste trabalho, a adoção desse padrão arquitetural

de *software* foi crucial. Ele permite a criação de um código bem estruturado em camadas, o que facilita a manutenção, testes e flexibilidade do sistema. Além disso, esse padrão tem como objetivo criar sistemas independentes de *frameworks*, o que facilita eventuais mudanças de tecnologias futuramente, se necessário.

### 2.5.1 Entidades

**Martin** enfatiza que as entidades do sistema encapsulam as regras críticas de negócios para toda a empresa. Independentemente da estrutura de dados utilizada para representar as entidades, elas devem ser facilmente reutilizadas em outras aplicações diferentes existentes no empreendimento. Em qualquer aplicação sendo desenvolvida, nenhuma mudança nas camadas externas devem afetar a camada de entidades.

### 2.5.2 Casos de Uso

Nesta camada, o software contém as regras de negócios específicas para aquela aplicação. Com a implementação dos casos de uso do sistema, ela organiza o fluxo de dados entre as entidades e as regras críticas de negócios, a fim de atingir um determinado objetivo. **Martin** reitera que esta camada também é isolada de qualquer camada externa, não dependendo de mudanças de bancos de dados, de interfaces do usuário ou até mesmo de *frameworks*. Isto se dá pelo fato de os casos de uso dependerem exclusivamente de interfaces bem definidas e não da implementação das mesmas.

### 2.5.3 Adaptadores de Interface

Esta é a camada responsável por converter os dados entre a camada de lógica de negócios e os detalhes externos, como *frameworks*, interfaces de usuário, tecnologias de persistência de dados. Esses adaptadores garantem a comunicação eficiente entre as camadas e podem ser divididos em:

- **Controladores:** desempenham um papel crucial como mediadores, conectando as ações dos usuários à lógica de negócios do sistema. Em uma aplicação web, esses adaptadores assumem a responsabilidade de receber comunicações HTTP, extrair dados relevantes para a aplicação e, fundamentalmente, iniciar a execução dos casos de uso específicos destinados a cada ação. Sua função é vital, proporcionando uma interface eficiente entre a interação do usuário e a operação fluida da aplicação.
- **Presenters e Gateways de Interface:** os *presenters* e *gateways* de interface desempenham papéis essenciais, pois facilitam a transição suave entre a camada de lógica de negócios e os detalhes externos, como interfaces de usuário e persistência em bancos de dados.

Enquanto os *presenters* dedicam-se a formatar os dados de maneira adequada para exibição, os *gateways* de interface assumem responsabilidade de garantir a persistência dos dados. Essa interação sinérgica entre *presenters* e *gateways* é fundamental para assegurar a compreensibilidade e a durabilidade dos dados em diferentes contextos dentro do sistema.

#### **2.5.4 *Drivers e Frameworks***

Esta é a camada mais externa do diagrama proposto, sendo constituída basicamente pelos *frameworks* e ferramentas de acesso aos diversos sistemas de gerenciamento de bancos de dados existentes. Esta camada tem como objetivo principal se comunicar com as camadas mais internas e manter os detalhes de tecnologia desacoplados do resto do código. [Martin](#) afirma que a tecnologia utilizada é um detalhe que deve ser mantido sem dependência com o código principal, evitando possíveis danos ao mesmo.

# 3

## Proposta

Atualmente, a atenção voltada para a privacidade na Internet tem crescido significativamente, dada a enorme quantidade de informações pessoais sendo coletadas, armazenadas, transmitidas e publicadas *online* [7]. As preocupações com a privacidade aumentaram devido a casos de violações de dados<sup>1</sup>, uso indevido de informações pessoais e falta de transparência por parte das empresas.

Devido aos crescentes problemas citados anteriormente e à sensibilidade em torno do uso inadequado de dados pessoais, faz-se necessário que os sites implementem as normas da Lei Geral de Proteção de Dados (LGPD) em suas páginas. A lei foi estabelecida com o intuito de garantir que as informações das pessoas naturais sejam tratadas de forma ética, transparente e segura [5].

Entretanto, a implementação das normas estabelecidas tornou-se cada vez mais complexa e desafiadora [32], dificultando para as empresas que possuem aplicações *web* garantir a conformidade com a lei. Isso requer uma análise cuidadosa por parte de especialistas legais para assegurar que os *sites* estejam em conformidade com a legislação.

Este trabalho foi desenvolvido visando automatizar o processo de análise de uma página da *web* que deveria ser feito por um (ou mais) especialista humano, de forma manual. É crucial não apenas garantir a conformidade com a lei, mas agilizar e simplificar o processo de análise, permitindo que as empresas identifiquem e corrijam potenciais não conformidades com a lei de forma eficiente e eficaz.

Neste capítulo, detalharemos o desenvolvimento da aplicação (em nível de prova de conceito) para avaliar a conformidade de páginas *web* em relação aos itens: segurança, privacidade, dados pessoais e consentimento, detalhados na LGPD.

---

<sup>1</sup><https://fia.com.br/blog/vazamento-de-dados/>

### 3.1 Arquitetura Geral do Projeto

A construção da aplicação depende de um conjunto de etapas, como apresentado na Figura 3.1 de forma simplificada. Ao longo deste capítulo, descreveremos cada uma dessas etapas.

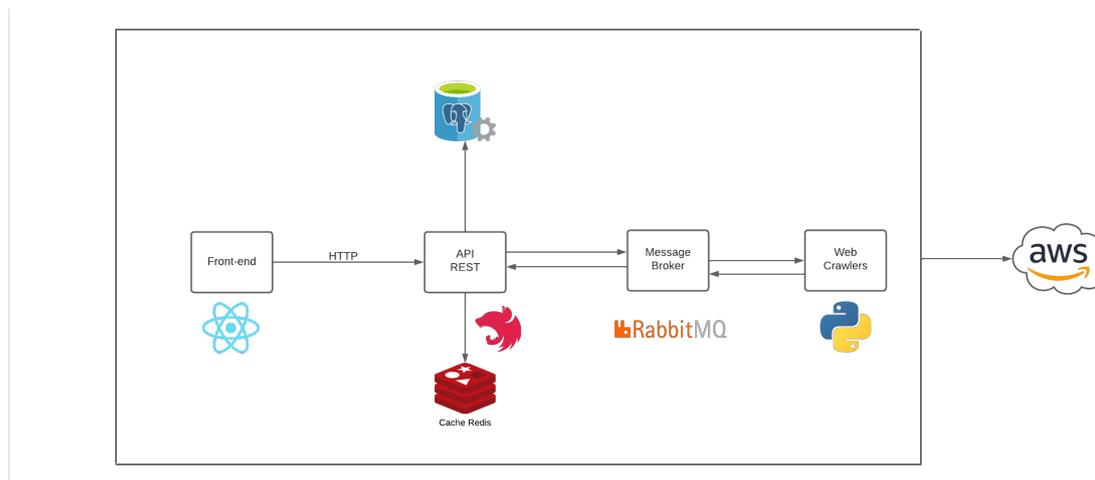


Figura 3.1: Arquitetura para o desenvolvimento da aplicação (Fonte: Elaborada pelo autor).

De forma resumida, para o *Front-end*, foram construídas páginas que irão interagir com os usuários da aplicação. Nesta camada, o endereço (a *Uniform Resource Locator - URL*) da página a ser analisada deve ser fornecido pelo usuário, a qual será processada em busca de inconformidades em relação à LGPD e, em seguida, será gerado um relatório contendo recomendações de ajustes.

Para a implementação do *Back-end* foi desenvolvido um servidor HTTP (vide Capítulo 2) para se comunicar diretamente com o *Front-end* da aplicação e iniciar o processamento da página informada pelo usuário. Além disso, caso a página já tenha sido processada anteriormente, foi incorporado um banco de dados de *cache* que armazena essas informações por 10 dias (esse número de dias pode ser configurado na ferramenta), evitando processamentos repetidos e, possivelmente, desnecessários.

Na etapa subsequente, foi construído um serviço de mensageria no formato de fila que receberá informações sobre os *sites* a serem processados pela aplicação. Este serviço possibilita que as páginas analisadas sejam processadas de forma assíncrona, evitando a sobrecarga do servidor *Back-end* por múltiplas requisições.

Na última camada da aplicação, os *web crawlers* (vide Capítulo 2) foram construídos com o intuito de extrair a marcação *HTML*<sup>2</sup> da página *web* e analisar seu conteúdo através do *Web Scraping*, buscando inconsistências em relação à LGPD. Por fim, toda a aplicação foi hospedada em ambiente de produção no serviço *Amazon Elastic Cloud Computing (EC2)*<sup>3</sup>.

<sup>2</sup><https://developer.mozilla.org/pt-BR/docs/Web/HTML>

<sup>3</sup><https://aws.amazon.com/pt/ec2/>

## 3.2 Front-End

Nesta seção, apresentaremos em detalhes a criação das páginas *web* da aplicação, destinadas à interação com o usuário. Abordaremos as tecnologias empregadas no desenvolvimento, o processo de criação do design das telas e os objetivos específicos de cada uma delas.

### 3.2.1 Tecnologias utilizadas

A linguagem de programação escolhida foi o Typescript<sup>4</sup>, devido aos seus recursos, como tipagem estática e interfaces, não existentes no Javascript<sup>5</sup> convencional, tornando mais fácil a detecção e prevenção de erros durante a fase de desenvolvimento.

Utilizamos, também, a biblioteca ReactJS<sup>6</sup>. Criada pelo Facebook, ela é uma das bibliotecas de Javascript mais populares e utilizadas no mundo inteiro em aplicações atuais. Além disso, possui a sua versão para desenvolvimento de aplicações *mobile* multiplataforma, o React Native<sup>7</sup>, sendo uma opção bastante prática para a criação de um aplicativo de celular para a evolução do escopo do trabalho atual (trabalhos futuros).

### 3.2.2 Telas do sistema

Para a prototipação das telas do sistema, foi utilizada a ferramenta de *design* Figma<sup>8</sup>, amplamente reconhecida e usada pela comunidade global de *designers* de interfaces de usuário. Sua principal vantagem é a sua capacidade de permitir que múltiplos usuários trabalhem em um mesmo projeto simultaneamente. Além disso, oferece uma ampla gama de recursos, incluindo a criação de *wireframes*, protótipos interativos, *design* de interfaces responsivas etc.

Na Figura 3.2, temos a página inicial do sistema. Nela, o usuário poderá visualizar os itens da LGPD que estão sendo analisados pela ferramenta no momento atual, junto com uma breve descrição de cada item considerado durante a fase de processamento. Além disso, o usuário poderá navegar pela barra de navegação ou iniciar um novo processamento clicando no botão "Testar agora".

Na parte inferior da Figura 3.2, está o rodapé do *website*, que apresenta ícones que, quando selecionados, permitem que o usuário entre em contato com o autor da pesquisa através de outras plataformas digitais.

---

<sup>4</sup><https://www.typescriptlang.org/>

<sup>5</sup><https://developer.mozilla.org/pt-BR/docs/Web/JavaScript>

<sup>6</sup><https://react.dev/>

<sup>7</sup><https://reactnative.dev/>

<sup>8</sup><https://www.figma.com/>



Figura 3.2: Página inicial da aplicação (Fonte: Elaborada pelo autor).

Na Figura 3.3, temos a página que detalha os critérios avaliados pela ferramenta em conformidade com a LGPD. Nesta página, os usuários podem compreender detalhadamente o significado de cada critério e aprender um pouco mais sobre a importância deles na garantia dos direitos digitais das pessoas naturais.

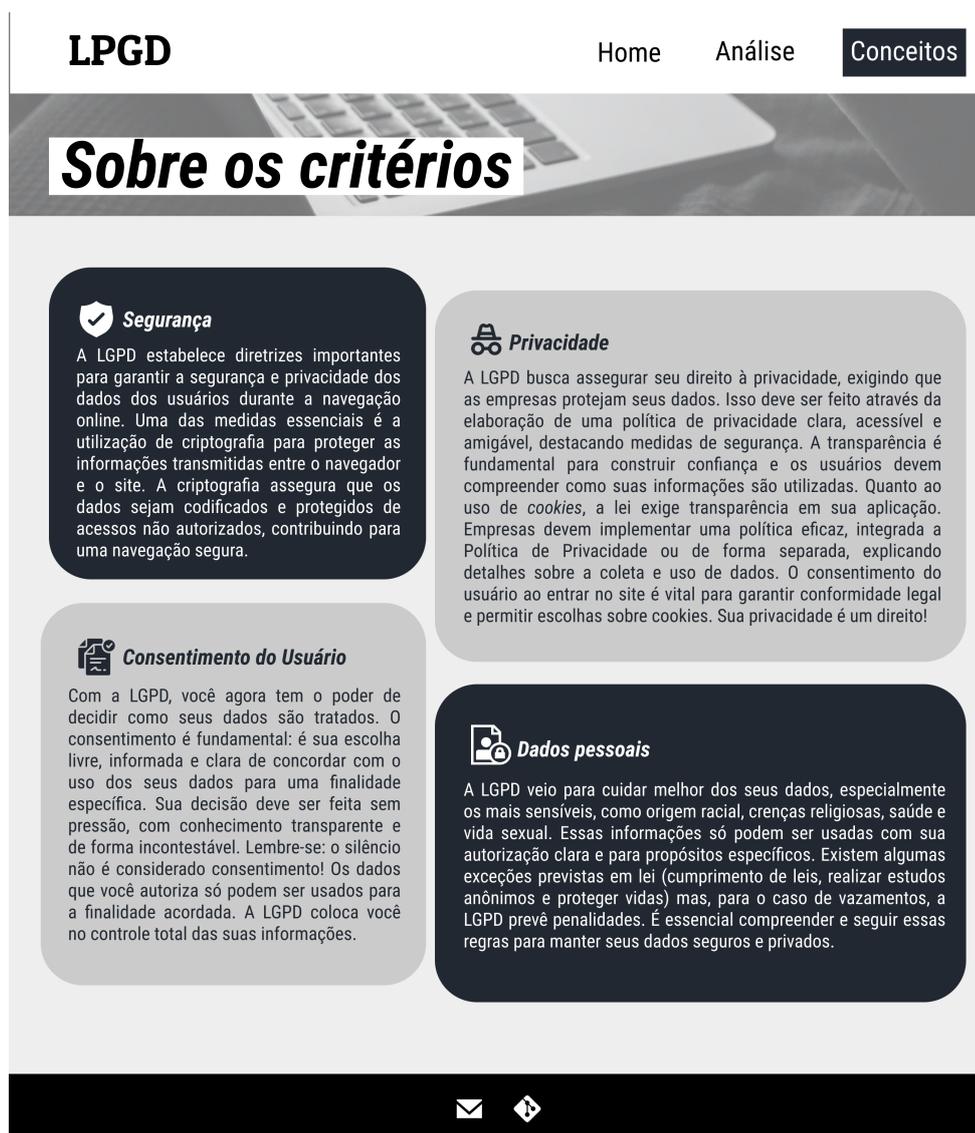


Figura 3.3: Página sobre os critérios abordados (Fonte: Elaborada pelo autor).

Na Figura 3.4, temos a tela em que o usuário fornece a *URL* do *website* que será analisado pela ferramenta. Após preencher o campo de entrada e clicar no botão "Testar agora", o usuário será redirecionado para a página de resultados. Se o processamento estiver em andamento, uma mensagem será exibida.

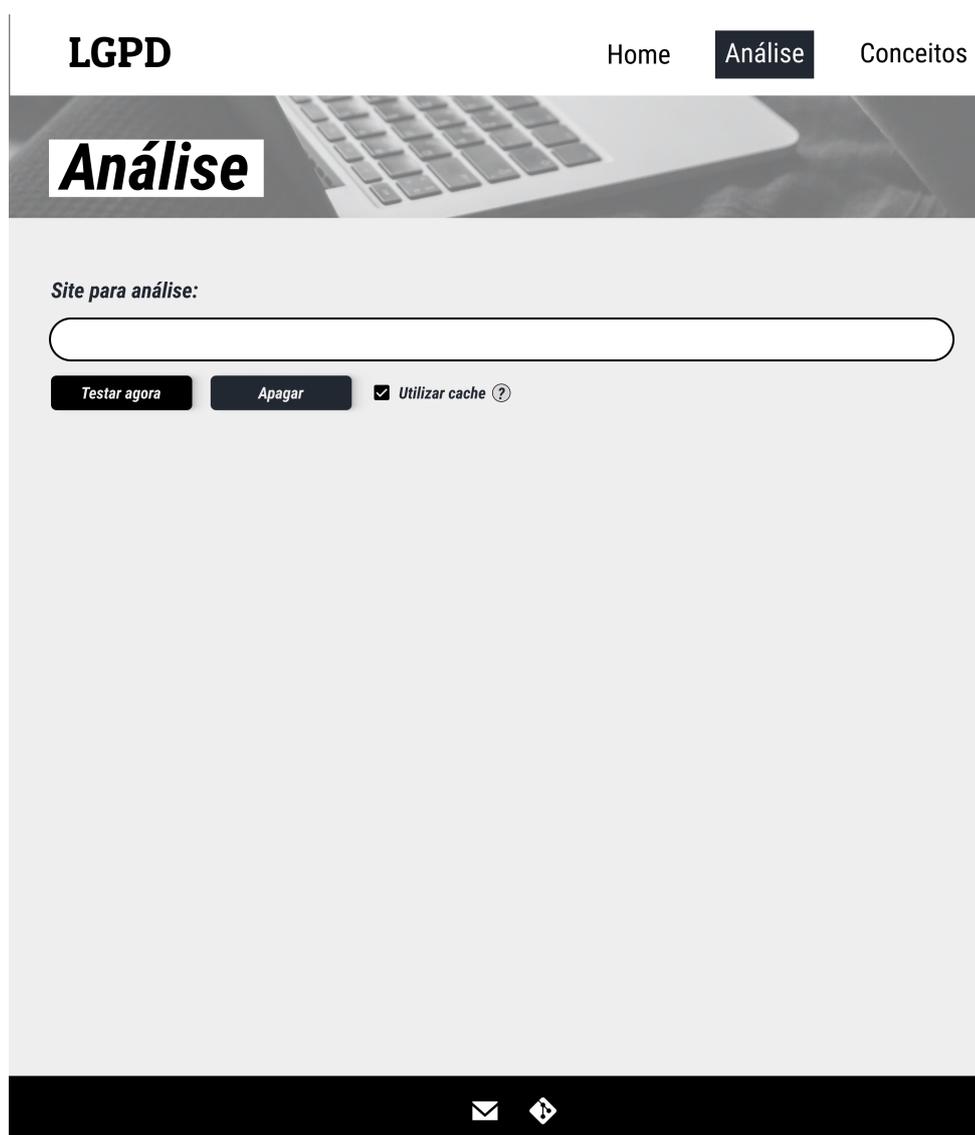


Figura 3.4: Página de início do processamento (Fonte: Elaborada pelo autor).

Na Figura 3.5, apresenta-se a tela dos resultados obtidos no processamento da página *web*. Nela, o usuário poderá visualizar as informações extraídas e processadas pela ferramenta por meio de uma tabela que indica se um determinado critério foi identificado ou não.



Figura 3.5: Página de resultados do processamento (Fonte: Elaborada pelo autor).

Por fim, na Figura 3.6, apresenta-se a tela das recomendações feitas pela aplicação. Após analisar o conteúdo extraído da página *web*, a ferramenta identifica e aponta as não conformidades com a LGPD e sugere ações necessárias para corrigir as falhas encontradas.

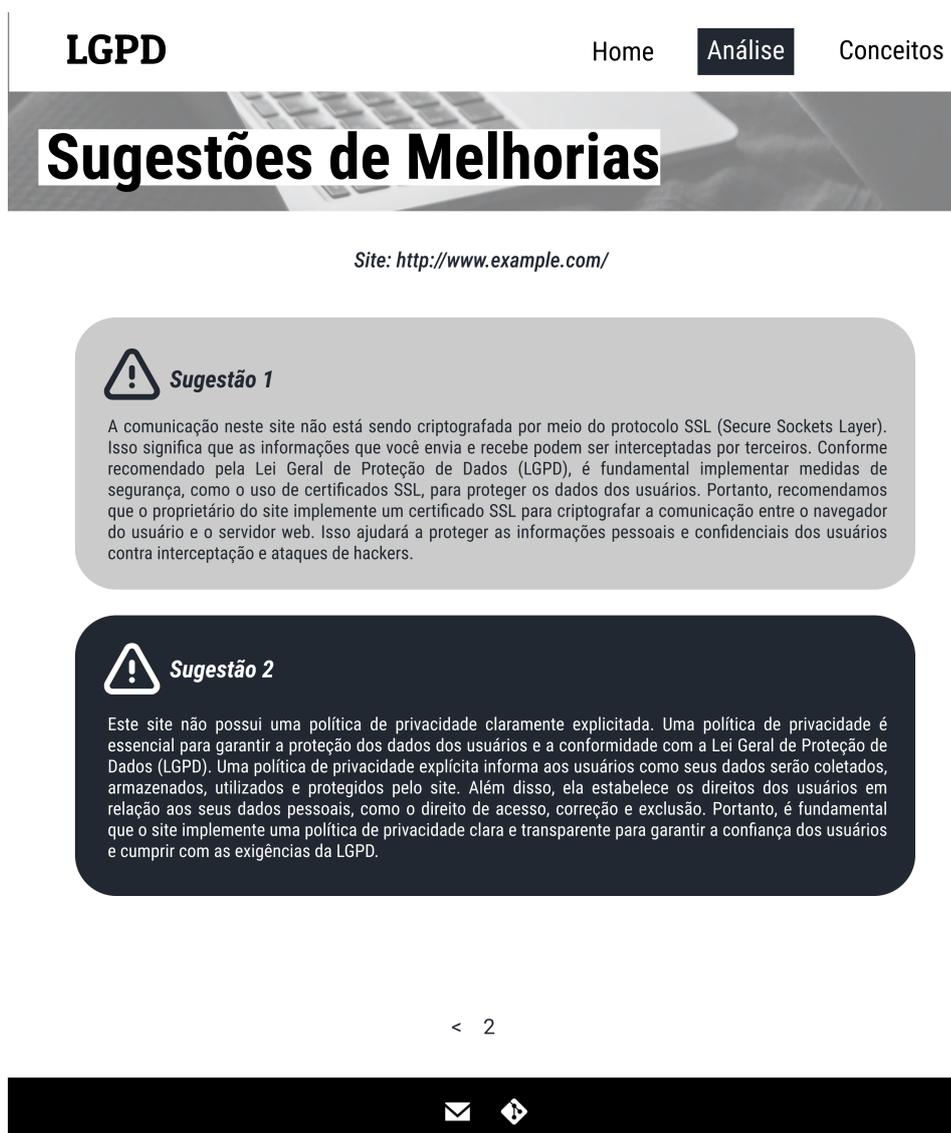


Figura 3.6: Página de resultados do processamento (Fonte: Elaborada pelo autor).

### 3.3 Back-end

Nesta seção, será descrito o desenvolvimento do serviço de *Back-end* para se comunicar com as páginas da aplicação e realizar o início da fase de processamento. Também serão explicitadas as tecnologias utilizadas na fase de desenvolvimento, a arquitetura de *software* utilizada, a comunicação com o serviço de mensageria e a comunicação com o banco de dados principal e o de *cache* e suas finalidades.

### 3.3.1 Tecnologias utilizadas

A linguagem de programação escolhida foi o Typescript<sup>9</sup>. Além das vantagens citadas na seção de *Front-end*, ela foi escolhida pela existência de *frameworks back-end* escritos na linguagem, bastante atuais e que apoiam o desenvolvimento de aplicações. Com isso, visamos agilidade e controle no desenvolvimento da aplicação.

Arelado à linguagem de programação, foi escolhido o *framework* Nest.js<sup>10</sup>, para a criação de aplicações HTTP (Hypertext Transfer Protocol) do lado do servidor. Essa escolha foi feita com o objetivo de desenvolver um serviço escalável e eficiente com uma tecnologia de documentação vasta e objetiva.

Para estabelecer uma comunicação entre o *Back-end* e o serviço de mensageria, foi utilizada a biblioteca AMQP Lib<sup>11</sup>, a qual utiliza o protocolo de comunicação assíncrona AMQP (*Advanced Message Queuing Protocol*).

Foram utilizados dois tipos de bancos de dados para o armazenamento das informações coletadas e processadas: o banco principal é PostgreSQL<sup>12</sup>, que tem como função armazenar todos os dados processados por tempo indeterminado, permitindo uma análise posterior dos mesmos. O segundo banco de dados utilizado foi Redis<sup>13</sup>, cujo objetivo é criar um sistema de *cache* e impedir que uma página já processada, nos 10 dias anteriores, seja processada novamente.

### 3.3.2 Arquitetura de *software* utilizada

Com o intuito de estruturar o projeto do *Back-end*, optamos pela implementação da Arquitetura Limpa [19]. A escolha desse padrão arquitetural de *software* fundamenta-se na capacidade de desenvolver um sistema flexível, desacoplado de tecnologias específicas, e facilmente testável. Dessa forma, caso seja necessário incorporar novas tecnologias ou remover tecnologias existentes, tais modificações podem ser realizadas com facilidade, sem impactar o núcleo do código da aplicação.

### 3.3.3 Entidades do sistema

Para o armazenamento dos dados de processamento (em nível de prova de conceito), foi estabelecido que era necessário apenas a modelagem de uma única entidade no banco de dados PostgreSQL<sup>10</sup>.

A Figura 3.7 exhibe a estrutura da tabela criada no banco de dados. A coluna "id" corresponde à chave primária da tabela, enquanto "siteUrl" armazena o endereço do *site*

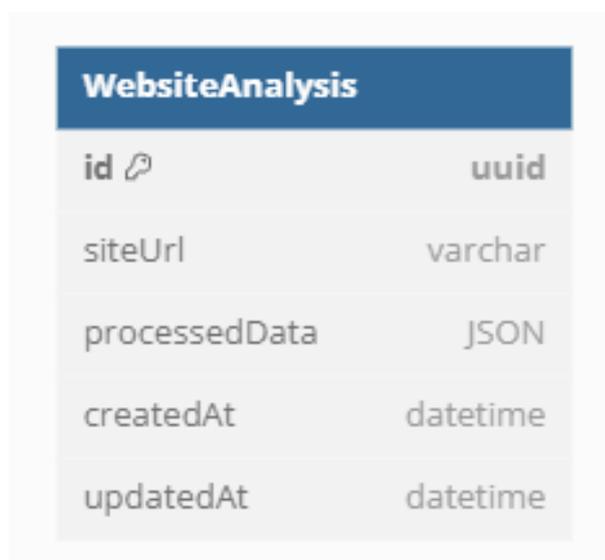
<sup>9</sup><https://www.typescriptlang.org/>

<sup>10</sup><https://nestjs.com/>

<sup>11</sup><https://github.com/amqp-node/amqplib>

<sup>12</sup><https://www.postgresql.org/>

<sup>13</sup><https://redis.io/>



WebsiteAnalysis	
id 	uuid
siteUrl	varchar
processedData	JSON
createdAt	datetime
updatedAt	datetime

Figura 3.7: Entidade de análise de um site (Fonte: Elaborada pelo autor).

hospedado na *Internet* e "processedData" guarda o resultado do processamento realizado pelos *web crawlers* no formato JSON. As colunas "createdAt" e "updatedAt" são destinadas ao registro das datas de criação e edição da entidade, permitindo o controle do histórico de alterações.

### 3.3.4 Cache

Em um cenário real, é comum que os usuários forneçam o mesmo endereço de uma página para ser processada e obter os resultados desejados. Contudo, isso pode resultar em execuções desnecessárias e repetitivas. A fim de evitar esse problema, foi implementado um mecanismo de *cache* que armazena as informações de processamento de uma página específica por até 10 dias. Dessa forma, o sistema pode reconhecer se uma página já foi processada recentemente e evitar a repetição do processamento, economizando recursos e garantindo uma melhor eficiência no uso da ferramenta.

Se o usuário preferir executar um novo processamento sem recorrer ao mecanismo de *cache* do sistema, ele terá a opção de marcar uma caixa de seleção que indica o desejo de processar o *site* novamente, como observado na Figura 3.4. No entanto, será alertado de que os resultados podem ser obtidos de maneira mais lenta devido ao processamento adicional.

A escolha do banco de dados Redis<sup>14</sup> para a implementação do mecanismo de *cache* se deu pelo fato de que ele armazena seus dados primariamente na memória RAM, sendo amplamente utilizado em aplicações que necessitam recuperar dados com baixa latência e alto desempenho. Além disso, o Redis oferece um mecanismo nativo que permite que o desenvolvedor defina um tempo de expiração para os dados armazenados. Esse recurso foi usado para o processo de cache mencionado anteriormente.

<sup>14</sup><https://redis.io/>

Na Figura 3.8, é possível ver um exemplo prático do armazenamento de dados com a definição de um tempo de expiração de 10 segundos.

```
redis> SET mykey "Hello"
"OK"
redis> EXPIRE mykey 10
(integer) 1
redis> TTL mykey
(integer) 10
```

Figura 3.8: Exemplo do mecanismo de expiração de dados (Fonte: red).

### 3.4 Serviço de mensageria

Considerando a preparação do sistema para cenários práticos, foi adotado um serviço de processamento de dados baseado em filas. Nesse modelo, os processamentos são realizados individualmente, seguindo a ordem de chegada na fila. A tecnologia selecionada para viabilizar essa estrutura de fluxo de dados foi o RabbitMQ<sup>15</sup>, devido à sua facilidade de uso e à capacidade de integração com diversas linguagens de programação atuais.

Na Figura 3.9, está a fila "*process\_website*" que foi criada para receber as mensagens para processamento.

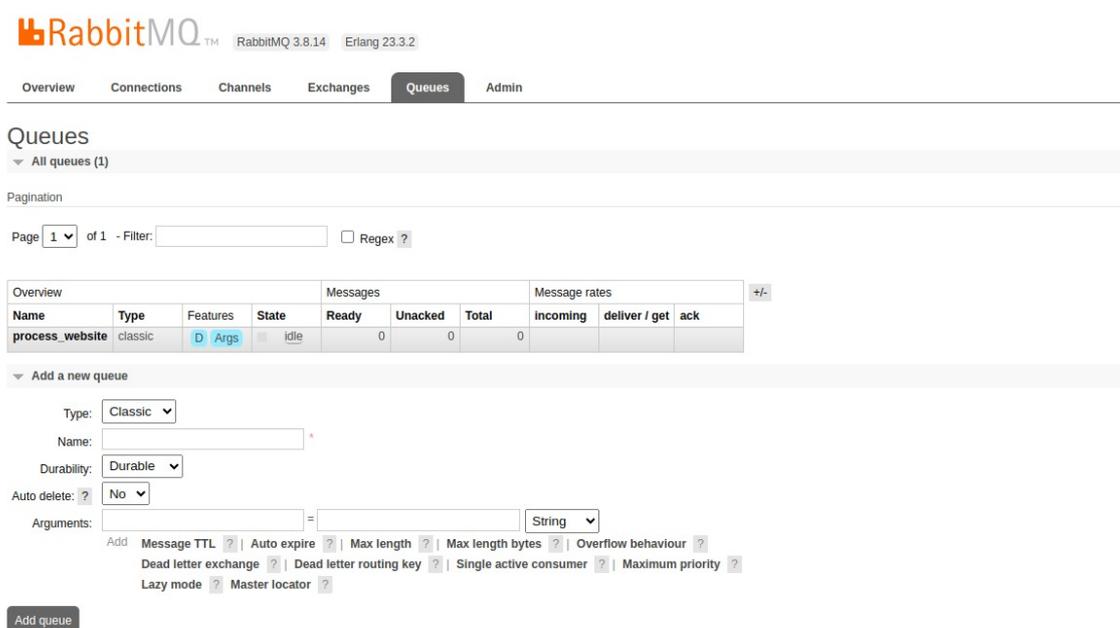


Figura 3.9: Imagem da interface do RabbitMQ (Fonte: Elaborada pelo autor).

<sup>15</sup><https://www.rabbitmq.com/>

## 3.5 Web Crawlers

A implementação dos *Web Crawlers*, responsáveis por extrair informações de páginas da *web*, foi feita utilizando a linguagem de programação Python<sup>16</sup> atrelada à biblioteca BeautifulSoup<sup>17</sup>. A escolha dessas tecnologias se deu pela combinação de facilidade de uso da linguagem e da eficiência e praticidade oferecidas pela biblioteca para realizar operações de coleta de dados. Além disso, foi utilizada a biblioteca Selenium<sup>18</sup> para a obtenção do código *HTML* das páginas.

### 3.5.1 Critérios avaliados

Dada a complexidade e amplitude da lei, que estabelece uma série de diretrizes a serem seguidas pelas empresas que operam aplicações digitais, optamos por selecionar apenas alguns critérios-chave para análise em um nível inicial de prova de conceito. Para cada critério estabelecido, foi desenvolvido um *crawler* que processa todos os itens relacionados a esse critério. Estes critérios incluem:

- Segurança (Artigo 6º, Inciso 7)
  - Comunicação criptografada: verifica se o *site* está utilizando um certificado *SSL* para a criptografia do tráfego de dados.
- Consentimento (Artigo 7º, Inciso 5)
  - Existência de termos de uso: analisa se o *site* possui alguma informação sobre os termos de uso da plataforma exposta na página.
  - Consentimento aos termos de uso: verifica se a página apresenta, de maneira clara e transparente, a necessidade de o usuário ler os termos de uso e concordar com as cláusulas estabelecidas.
  - Opção para aceitar os termos de uso: analisa se a página possui uma caixa de seleção para aceitar os termos de uso da plataforma antes de prosseguir com uma ação.
- Privacidade (Artigo 2º, Inciso 1)
  - Política de privacidade: verifica se a página possui alguma informação sobre a política de privacidade adotada.
  - *Cookies*: avalia se o *site* faz uso de *cookies* e se comunica, de forma transparente com o usuário, sobre os detalhes desta prática.

---

<sup>16</sup><https://www.python.org/>

<sup>17</sup><https://www.crummy.com/software/BeautifulSoup/>

<sup>18</sup><https://www.selenium.dev/>

- Crianças e adolescentes (Artigo 14º)
  - Informações sobre menores de idade: verifica se a página apresenta alguma informação sobre menores de idade, sem os devidos cuidados.
- Dados pessoais (Artigo 1º)
  - CPF: verifica se a página expõe o CPF de uma pessoa natural em seu conteúdo.
  - CEP: analisa se a página expõe a informação de um CEP em seu conteúdo.
  - *E-mail*: verifica se a página expõe um e-mail em seu conteúdo.
  - Número de telefone: analisa se a página expõe um número de telefone em seu conteúdo.
  - Endereço: analisa se a página expõe um endereço em seu conteúdo.
  - CNPJ: analisa se a página expõe um CNPJ em seu conteúdo.

É importante ressaltar que a LGPD, em seu Artigo 1º, estabelece como principal objetivo a proteção da pessoa física no tratamento de dados pessoais. Nesse contexto, informações como o CNPJ não seriam de interesse para análise. No entanto, é fundamental que o responsável pelo *site* esteja atento, pois algumas informações podem identificar uma pessoa física mesmo que não seja imediatamente óbvio [2]. Dessa forma, a ferramenta também detecta essas informações, visando alertar sobre não conformidades com a LGPD.

### 3.5.2 Desenvolvimento

O desenvolvimento foi segmentado em duas partes distintas. A primeira etapa envolve o desenvolvimento de cada *crawler* específico, responsável por uma tarefa designada. Esses *crawlers* são implementados como classes abstratas, cada uma obrigada a implementar o método "*crawl*" (observado na Figura 3.10), que encapsula a execução da tarefa atribuída e retorna os resultados do processamento realizado pelo mesmo.

Caso haja interesse em expandir o sistema para incluir mais critérios da LGPD, será necessário apenas criar novos *crawlers* que se encarregarão do processamento dos dados de acordo com esses critérios.

```
def crawl(self):
    self.has_terms_of_use()
    self.has_accept_string()
    self.has_checkbox()
    return self.results
```

Figura 3.10: Método que executa as tarefas do *crawler* (Fonte: Elaborada pelo autor).

Na Figura 3.11, observa-se uma tarefa responsável por verificar se a página do *site* contém uma caixa de seleção que informa ao usuário sobre a necessidade de aceitar os termos de uso da aplicação para prosseguir com a ação desejada.

```
def has_checkbox(self):
    checkboxes = self.soup.find_all('input', {'type': 'checkbox'})

    for checkbox in checkboxes:
        attrs = checkbox.attrs
        for attr in attrs:
            if isinstance(checkbox[attr], str):
                if verify_string_intext(strings=['termos', 'terms'], text=checkbox[attr]):
                    self.results['has_checkbox'] = True
                    return

    self.results['has_checkbox'] = False
```

Figura 3.11: Exemplo de uma tarefa de um *crawler* (Fonte: Elaborada pelo autor).

Na segunda etapa, foi elaborada uma classe encarregada de atuar como intermediária entre os *crawlers* e as mensagens recebidas pela fila. Além disso, essa classe é responsável por extrair o código *HTML* da página analisada e realizar a remoção das *tags* desnecessárias para a aplicação. A Figura 3.12 apresenta a instanciação dos *crawlers* na classe.

```
self.crawlers = [
    {"key": "cookies", "crawler": PrivacyCrawler(soup=self.soup)},
    {"key": "ssl", "crawler": SSLCrawler(url=self.url)},
    {"key": "personal_data",
     "crawler": PersonalDataCrawler(soup=self.soup)},
    {"key": "consent", "crawler": ConsentCrawler(soup=self.soup)},
    {"key": "child", "crawler": ChildCrawler(soup=self.soup)}
]
```

Figura 3.12: Instâncias dos *crawlers* (Fonte: Elaborada pelo autor).

Na Figura 3.13, está o método que inicia o processamento da página executando um *crawler* por vez e salvando o resultado da análise.

```
def verify(self):
    print(">> INICIANDO CRAWLERS")

    for crawler_info in self.crawlers:
        key, crawler = crawler_info.values()
        self.results[key] = crawler.crawl()

    return self.results
```

Figura 3.13: Execução dos *crawlers* (Fonte: Elaborada pelo autor).

Por fim, a camada dos *crawlers* é inicializada por meio de um *script* que utiliza a classe desenvolvida na segunda etapa do processo de desenvolvimento. Dentro desse *script*, é implementado um consumidor (Figura 3.14) da fila do serviço de mensageria, que permanece aguardando o envio de novas mensagens. Quando uma mensagem é recebida, uma função de retorno é acionada, dando início ao processo de processamento. Quando o processamento é finalizado, o resultado é salvo nos bancos de dados.

```
def callback(ch, method, properties, body):
    try:
        message = json.loads(body.decode('utf-8'))
        if 'url' in message and 'testId' in message:
            verifier = Verifier(url=message['url'])
            results = verifier.verify()
            test = Test.get(Test.id == message['testId'])
            if test:
                test.processedData = results
                test.save()

            redis_test = json.loads(db_redis.get(f"site:{message['url']}"))
            if redis_test:
                redis_test['processedData'] = results
                db_redis.set(f"site:{message['url']}", json.dumps(
                    redis_test), kepttl=True)

        except Exception as e:
            print(str(e))

        ch.basic_ack(delivery_tag=method.delivery_tag)

    print("Aguardando mensagens...")
    channel.basic_consume(queue=queue_name, on_message_callback=callback)
    channel.start_consuming()
```

Figura 3.14: Implementação do consumidor (Fonte: Elaborada pelo autor).

## 3.6 Ambiente de produção

Por fim, a ferramenta foi hospedada em um ambiente de produção utilizando o serviço EC2<sup>19</sup> da Amazon. Para simplificar a implantação dos serviços nesse ambiente, recorreremos à ferramenta de virtualização de *containers* Docker<sup>20</sup>. Esta ferramenta automatiza a execução das tecnologias, baseando-se em um arquivo de configuração, e padroniza as versões das mesmas. A Amazon disponibiliza vários tipos de instâncias de máquinas para utilização. As configurações da máquina escolhida foram:

- Tipo de instância: *t2.medium*
- vCPUs: 2
- RAM (GIB): 4
- Volume da instância: Dinamicamente alocado

Vale ressaltar que este ambiente foi criado para a simulação de um ambiente de produção e validação da prova de conceito. No caso de uma aplicação que seria utilizada por muitos usuários diariamente, seria necessário hospedar a ferramenta em um ambiente mais robusto.

## 3.7 Objetivo geral

A proposta apresentada neste capítulo tem como objetivo geral automatizar a análise manual de sites no que diz respeito ao cumprimento das cláusulas estabelecidas pela LGPD.

## 3.8 Objetivos específicos

Com essa abordagem, almeja-se reduzir o tempo necessário para realizar essas análises, substituindo processos manuais por uma solução automatizada. A ferramenta desenvolvida aumentará a eficiência do processo e, também, possibilitará uma análise mais consistente, abrangente e padronizada. Isso permitirá que os proprietários dos *sites* identifiquem não conformidades em relação aos critérios implementados.

Ao automatizar a verificação das cláusulas da LGPD, pretende-se auxiliar os proprietários de aplicações *web* e facilitar o trabalho dos profissionais responsáveis pela auditoria legal, proporcionando-lhes uma ferramenta para identificar áreas de não conformidade e recomendar medidas corretivas necessárias aos seus clientes.

---

<sup>19</sup><https://aws.amazon.com/pt/ec2/>

<sup>20</sup><https://www.docker.com/>

Além disso, busca-se avaliar o nível de aceitação da ferramenta junto ao público-alvo selecionado para a realização de um experimento. Para isso, foi utilizado um formulário online que permitirá a coleta de opiniões sobre a usabilidade e relevância da ferramenta. Essa abordagem proporcionará conclusões valiosas sobre como a ferramenta é percebida pelos usuários e identificará áreas para melhorias, contribuindo assim para o aprimoramento contínuo do sistema.



# Metodologia

Neste capítulo, serão descritas as estratégias adotadas ao longo da condução deste trabalho, contendo a definição das abordagens de pesquisa e os procedimentos específicos utilizados para coleta e análise dos dados. Será apresentado o contexto em que a pesquisa foi desenvolvida, incluindo a justificativa das escolhas metodológicas de acordo com os objetivos propostos. Além disso, serão detalhados os instrumentos empregados na coleta de informações, bem como os critérios de seleção dos participantes.

A análise manual de um *site* consiste na avaliação conforme os critérios da LGPD definidos no escopo deste trabalho. Essa avaliação foi realizada por meio de um formulário que inclui itens relacionados aos critérios selecionados. Por outro lado, a análise automatizada refere-se à mesma avaliação, mas realizada por meio da ferramenta desenvolvida.

## 4.1 Etapas da pesquisa

Com os objetivos da pesquisa definidos no capítulo anterior (vide Subcapítulo 3.7), fez-se necessário adotar uma estratégia para alcançá-los. Entre as estratégias selecionadas para a implementação da pesquisa, destacam-se:

- Seleção do público-alvo;
- Formulário de análise manual;
- Definição do experimento;
- Pesquisa de satisfação;
- Métricas de interesse.

## 4.2 Seleção do público-alvo

Durante esta fase, foi crucial estabelecer um número preciso de participantes para validar a solução implementada na ferramenta. Este estudo tem como objetivo atingir dois tipos de indivíduos: leigos, sem ou com escasso conhecimento sobre a Lei Geral de Proteção de Dados (LGPD), e especialistas em legislação, incluindo profissionais do direito.

O objetivo de atingir dois perfis distintos de usuários reside na intenção de compreender os benefícios e limitações da solução automatizada para: (1) indivíduos com conhecimento limitado sobre o tema em questão, e (2) indivíduos que fazem uso cotidiano dos conceitos avaliados.

Dessa forma, foram selecionados 30 participantes, que foram solicitados a preencher um formulário para coletar informações sobre a ferramenta, após terem realizado uma análise manual por conta própria nas páginas específicas pré-selecionadas de alguns *sites*. Não foi exigido que os participantes fossem profissionais do direito, mas que informassem seu nível de conhecimento sobre o tema.

## 4.3 Formulário de análise manual

A fim de auxiliar a fase posterior da execução do experimento, foi implementado um formulário com perguntas sobre a análise manual feita pelo participante. Este formulário é dividido em 5 seções:

- Termo de Consentimento Livre e Esclarecido (TCLE);
- Informações do participante;
- Análise manual;
- Fim do experimento.

### 4.3.1 Termo de Consentimento Livre e Esclarecido

A primeira seção do formulário é dedicada ao Termo de Consentimento Livre e Esclarecido (TCLE). Nela, o participante será informado sobre todos os dados coletados pela pesquisa e como serão utilizados. Os pesquisadores garantem e se comprometem com o sigilo e a confidencialidade de todas as informações fornecidas no experimento.

Ademais, será esclarecido que a pesquisa não será remunerada, pois a participação é voluntária. Para prosseguir, é necessário confirmar que leu o termo de consentimento e concorda em participar do experimento. O participante será informado de que, mesmo após participar da pesquisa, caso se arrependa, poderá solicitar a exclusão de sua participação.

### 4.3.2 Informações do participante

O objetivo desta seção é compreender o perfil dos participantes do experimento. Dado que o público-alvo da pesquisa inclui tanto pessoas leigas no assunto quanto especialistas em legislação e direito, busca-se identificar o nível de conhecimento dos participantes sobre o tema. Essa informação também será importante para a pesquisa de satisfação que será feita ao final da execução do experimento. Esta parte do formulário contém 4 questões:

- Você já participou de alguma pesquisa relacionada à Lei Geral de Proteção de Dados?
- Você possui experiência ou trabalha na área jurídica?
- Em relação ao seu conhecimento jurídico, o quanto você está familiarizado com a Lei Geral de Proteção de Dados?
- Se você fosse julgar o nível de confiabilidade da sua análise manual de um site para identificar se ele fere a LGPD ou não, em uma escala de 1 a 5, qual seria o seu julgamento?

As duas primeiras questões exigem respostas "Sim" ou "Não". A penúltima questão oferece opções como "Não conheço nada", "Já ouvi falar sobre", "Já li sobre o assunto, mas não sou especialista" e "Sou especialista em LGPD". Por fim, a última pergunta é avaliada em uma escala de 1 a 5, onde 1 indica "nada confiável" e 5 "muito confiável".

### 4.3.3 Análise manual

Nesta etapa do formulário, o usuário estará realizando o experimento de análise manual de cada página selecionada, verificando se há alguma informação que possa não estar em conformidade com a LGPD. Foram criadas 12 questões para essa seção:

- O site é HTTPS? (Caso seja, aparecerá um cadeado do lado esquerdo do endereço do site)
- A página fornece informações sobre coleta de cookies?
- A página permite que você gerencie as permissões de coleta de cookies?
- A página fornece informações sobre a política de privacidade adotada?
- A página fornece informações sobre os termos de uso que os usuários devem estar de acordo para navegar?
- A página apresenta explicitamente um campo para que o usuário preencha que leu e concorda com os termos de uso?
- A página apresenta explicitamente algum dado referente a um CPF?

- A página apresenta explicitamente algum dado referente a um número telefone?
- A página apresenta explicitamente algum dado referente a um endereço de e-mail?
- A página apresenta explicitamente algum dado referente a um CEP?
- A página apresenta explicitamente algum dado referente a um endereço físico?
- A página possui alguma informação sobre como lida com o acesso de crianças e adolescentes?

Cada questão está relacionada a um item analisado pela ferramenta e será respondida com "Sim" ou "Não". Essas mesmas questões serão repetidas para cada página analisada individualmente.

#### 4.3.4 Fim do experimento

Por fim, a última seção contém apenas uma única questão, cujo objetivo é coletar o tempo total que o participante levou para concluir o experimento. Essa questão é crucial para a fase de análise dos dados coletados.

### 4.4 Definição do experimento

Para iniciar o experimento com os participantes, foi necessário estabelecer uma série de procedimentos. Esses passos são fundamentais para garantir a padronização do experimento e aumentar sua confiabilidade.

Inicialmente, todos os participantes foram submetidos a um formulário de análise manual em um ambiente virtual e controlado, através da plataforma Google Meet<sup>1</sup>, recebendo o suporte necessário da equipe responsável pela pesquisa.

Nesse ambiente, os participantes receberam um breve resumo sobre a LGPD e uma explicação sobre os critérios avaliados no estudo, conforme o escopo pré-estabelecido. Durante a análise manual das páginas, os voluntários puderam consultar a lei para esclarecer possíveis dúvidas.

Após todos os esclarecimentos sobre a pesquisa, o experimento foi iniciado. Cada participante recebeu 4 endereços de páginas de sites escolhidas pela equipe e foi instruído a analisar cada página separadamente, preenchendo o formulário correspondente. Cada participante realizou o experimento individualmente, enquanto a equipe auxiliava na marcação do tempo necessário para sua conclusão.

As páginas de *sites* escolhidas para o experimento foram:

---

<sup>1</sup><https://meet.google.com/>

- [httpforever.com](http://httpforever.com)
- [empresasweb.net/emp/lucas-buarque-de-araujo-barros-46620940](http://empresasweb.net/emp/lucas-buarque-de-araujo-barros-46620940)
- [dcarvalho.com.br/fale-conosco/contato](http://dcarvalho.com.br/fale-conosco/contato)
- [brpods.com/politica-de-privacidade/](http://brpods.com/politica-de-privacidade/)

Após cada participante concluir a análise manual e preencher o formulário, eles realizaram a mesma análise utilizando a ferramenta, auxiliados pelos organizadores da pesquisa, com o objetivo de comparar o trabalho manual com o automatizado. Em seguida, cada participante prosseguiu para a outra seção do formulário relacionada à pesquisa de satisfação sobre o experimento.

## 4.5 Pesquisa de satisfação

O desenvolvimento da pesquisa de satisfação foi feito por meio de um formulário e se baseou no modelo TAM (Technology Acceptance Model), cujo principal objetivo é mensurar o nível de aceitação dos usuários a respeito da tecnologia desenvolvida [16].

Este modelo é uma estrutura amplamente reconhecida para compreender a adoção de tecnologia, focando em dois principais determinantes: a percepção da utilidade e a facilidade de uso. Dessa forma, é possível identificar fatores que influenciam a aceitação da tecnologia e, assim, direcionar esforços para melhorar a experiência do usuário e garantir uma maior adesão à ferramenta.

Para coletar as respostas dos usuários em relação às questões do questionário, optou-se pela utilização da escala de Likert. Essa metodologia permite avaliar o grau de concordância dos usuários com as afirmações apresentadas na pesquisa, fornecendo uma medida quantitativa da percepção dos participantes [13]. Os níveis de afirmação adotados foram:

- Concordo totalmente;
- Concordo parcialmente;
- Nem concordo nem discordo;
- Discordo parcialmente;
- Discordo totalmente.

### 4.5.1 Percepção de utilidade

A percepção de utilidade refere-se à avaliação dos participantes da pesquisa sobre a relevância da ferramenta automatizada na análise contínua das páginas de *websites*, com o objetivo

de identificar possíveis não conformidades com a LGPD. Para esse construto, foram elaboradas as seguintes afirmações:

- A ferramenta de análise automática de sites me ajuda a identificar facilmente as áreas do meu site que não estão em conformidade com a LGPD.
- Eu acredito que usar a ferramenta de análise automática me ajudará a melhorar a conformidade do meu site com a LGPD.
- A ferramenta de análise automática me permite compreender melhor algumas exigências da LGPD em relação ao meu site.
- Usar a ferramenta de análise automática é benéfico para manter a conformidade contínua com a LGPD no meu site.
- Usar a ferramenta de análise automática ajuda a evitar possíveis penalidades e multas por não conformidades com a LGPD.
- Se eu fosse um profissional do direito e precisasse identificar não conformidades em páginas de *websites*, a ferramenta de análise automática me ajudaria neste trabalho.
- Se eu fosse um profissional do direito e especialista em LGPD, eu utilizaria a ferramenta no meu trabalho.
- A ferramenta de análise automática poderia facilmente substituir um especialista em direito e LGPD.
- A ferramenta de análise automática não seria um substituto para um especialista em direito e LGPD, mas serviria como uma segunda opinião, complementando a expertise humana na identificação de possíveis não conformidades com a legislação.

#### 4.5.2 Facilidade de uso

A facilidade de uso é a percepção dos participantes em relação à simplicidade de utilizar a ferramenta. Para esse construto, foram elaboradas as seguintes afirmações:

- É fácil aprender a usar a ferramenta de análise automática de sites para a LGPD.
- A interface da ferramenta de análise automática é intuitiva e fácil de navegar.
- A ferramenta fornece *feedback* imediato sobre os problemas de conformidade encontrados, o que facilita a correção.
- Eu posso usar a ferramenta facilmente sem a necessidade de treinamento extensivo ou suporte técnico.

- O uso da ferramenta torna-se fácil e acessível a todos devido ao fato de ser uma aplicação *web*, não exigindo instalação nem configuração.
- Eu não preciso ser um profissional de direito e especialista em LGPD para utilizar a ferramenta.
- Eu não preciso entender de programação e ciência da computação para utilizar a ferramenta.

## 4.6 Métricas de interesse

A fim de analisar o desempenho de cada participante no experimento, foram definidas duas métricas de interesse: tempo médio de conclusão e erro médio.

Para avaliar o tempo médio de conclusão, foram registrados os tempos de conclusão das análises manuais de cada participante, considerando que todos foram submetidos ao experimento em um ambiente controlado, livre de distrações e padronizado. Em seguida, foi calculada a média desses tempos coletados.

No cálculo do erro médio entre os participantes, dado que o experimento foi conduzido de forma controlada e as não conformidades em cada página foram previamente identificadas, as respostas de todos os participantes foram comparadas com as respostas corretas para determinar o erro de cada um. Posteriormente, foi calculada uma média desses erros.

# 5

## Resultados e Discussões

Neste capítulo, serão apresentados os resultados obtidos a partir da pesquisa realizada com 30 participantes voluntários, visando comparar a análise manual, feita pelos participantes, com a análise automatizada utilizando a ferramenta desenvolvida neste trabalho.

A análise, conforme descrita na Metodologia (Capítulo 4), se refere à conformidade de *websites* com os critérios da Lei Geral de Proteção de Dados (LGPD) abordados na proposta deste trabalho (vide Capítulo 3). Para tanto, consideraremos as seguintes métricas para a avaliação: tempo médio para conclusão de uma análise e erro médio na identificação de não conformidades.

### 5.1 Perfil dos participantes

Inicialmente, convém descrever os participantes da pesquisa com base nas respostas dos mesmos à seção "informações do participante" do formulário do experimento. Destacam-se características pertinentes que podem impactar a análise comparativa entre os métodos manual e automatizado de conformidade com a LGPD.

#### 5.1.1 Participação em Pesquisas Relacionadas à LGPD

A maioria dos participantes (90% ou 27) indicou não ter participado anteriormente de pesquisas relacionadas à Lei Geral de Proteção de Dados (LGPD). Apenas 10% (3) dos participantes responderam positivamente. Este resultado sugere que a experiência anterior com estudos sobre a LGPD é limitada entre os participantes deste estudo (vide Gráfico 5.1).

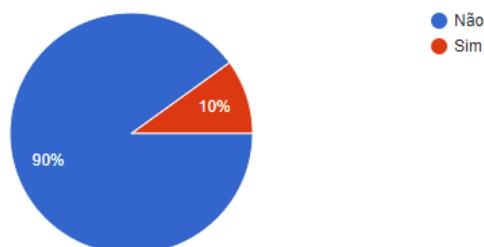


Figura 5.1: Gráfico obtido a partir das respostas da questão sobre participação em pesquisas anteriores relacionadas à LGPD (Fonte: elaborada pelo autor).

### 5.1.2 Experiência ou Trabalho na Área Jurídica

Uma parcela significativa dos participantes (73,3% ou 22) não possui experiência ou trabalho na área jurídica, enquanto (26,7% ou 8) afirmaram possuir essa experiência. Este dado é relevante para considerar o contexto e o nível de familiaridade dos participantes com conceitos jurídicos envolvidos na interpretação e aplicação da LGPD.

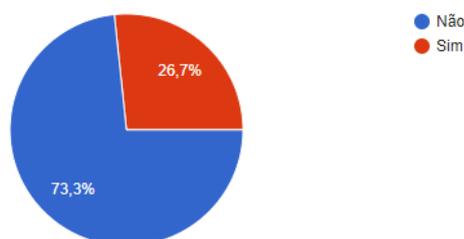


Figura 5.2: Gráfico das respostas sobre experiência jurídica (Fonte: elaborada pelo autor).

### 5.1.3 Nível de Familiaridade com a LGPD

Quando questionados sobre seu nível de familiaridade com a LGPD, os participantes relataram as seguintes informações:

- 53,3% (16) responderam que já leram sobre a LGPD, mas não se consideram especialistas;
- 40% (12) afirmaram já ter ouvido falar sobre a lei;
- 6,7% (2) indicaram não conhecer nada sobre a lei.

Esses resultados destacam uma distribuição variada no conhecimento prévio dos participantes sobre a legislação de proteção de dados, fator que pode influenciar na qualidade e na

confiabilidade de suas análises manuais, realizadas durante o estudo. Vale ressaltar que nenhum especialista em LGPD participou do experimento.

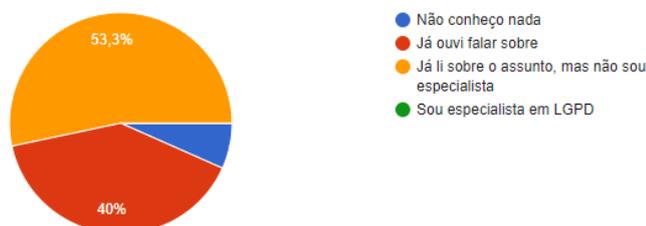


Figura 5.3: Gráfico das respostas sobre a familiaridade com a LGPD (Fonte: elaborada pelo autor).

#### 5.1.4 Julgamento do Nível de Confiabilidade na Análise Manual

Em relação o nível de confiabilidade das análises manuais dos participantes, para identificar possíveis não conformidades com a LGPD, as respostas dos participantes se distribuem da seguinte forma:

- 13,3% (4) atribuíram uma nota de 1;
- 23,3% (7) atribuíram uma nota de 2;
- 46,7% (14) atribuíram uma nota de 3;
- 13,3% (4) atribuíram uma nota de 4;
- 3,3% (1) atribuíram uma nota de 5.

Esses dados fornecem uma visão geral das percepções dos participantes quanto à confiabilidade de suas próprias análises (manuais), o que é crucial para comparar com os resultados obtidos pela ferramenta automatizada. É possível concluir que mais da metade dos participantes considera-se capaz, pelo menos minimamente, de identificar o que foi solicitado no experimento e emitir um parecer em relação à qualidade da realização da tarefa.

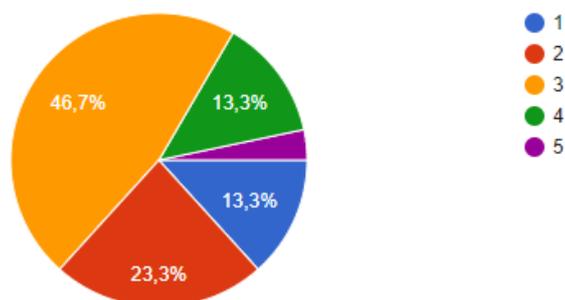


Figura 5.4: Gráfico das respostas sobre a confiabilidade da análise manual (Fonte: elaborada pelo autor).

## 5.2 Tempo de análise

Ao término de todas as análises manuais, registrou-se o tempo que cada participante levou para concluí-las. Assim, obteve-se:

- **Análise Manual:** O tempo médio registrado para que os participantes concluíssem a análise manual das 4 páginas especificadas foi de aproximadamente 9,94 minutos.
- **Análise Automatizada:** A ferramenta automatizada concluiu a análise das 4 páginas, no servidor de testes, em 26 segundos.

A comparação entre os tempos necessários para os métodos manual e automatizado sugere maior eficiência operacional da abordagem automatizada. Essa abordagem acelera o processo de verificação de conformidade, e possibilita a realização de tais verificações em larga escala. Assim, a ferramenta pode auxiliar um avaliador humano, no processo de verificação manual, apresentando sugestões de melhorias/correções ao encontrar não conformidades.

## 5.3 Precisão da Análise

Através das seções do formulário destinadas às análises manuais, foram coletados os erros nas análises (manuais e automática) das 4 páginas especificadas. Com o número de erros foi calculada uma média dos erros. Os resultados são apresentados abaixo: (Com o número de erros foram calculadas medidas estatísticas referentes a esta variável)

- **Análise Manual:** O erro médio identificado entre as respostas dos participantes ao analisar todas as páginas foi registrado como sendo aproximadamente 6,833, sendo a soma total de erros igual a 205.
- **Análise Automatizada:** A ferramenta automatizada apresentou uma precisão de 100% na análise das páginas pré-selecionadas.

Ao observar os resultados dos participantes, nota-se que, embora alguns tenham formação na área jurídica ou conhecimento prévio sobre a LGPD, a média de erro é elevada. Isso sugere a necessidade de capacitação e prática (assistida e independente) na verificação da LGPD (e, potencialmente, outras normas) em uma análise manual. Dessa forma, a ferramenta se torna uma opção viável para auxiliar na verificação manual.

É importante destacar que a ferramenta alcançou uma precisão de 100% em um ambiente controlado, onde os sites foram previamente selecionados e as não-conformidades com a LGPD mapeadas manualmente. No entanto, ao avaliar sites aleatórios, é possível que a precisão experimente uma leve redução.

## 5.4 Pesquisa de satisfação

Após a análise manual, foi solicitado aos participantes que avaliassem a ferramenta proposta e expressassem suas percepções. A seguir serão apresentados os resultados dessa pesquisa.

### 5.4.1 Percepção de utilidade da ferramenta

Com o intuito de avaliar a percepção dos participantes sobre a utilidade da ferramenta no mundo real, buscamos entender suas aplicações em situações práticas. Para isso, consultamos o nível de concordância em relação a algumas afirmações sobre a ferramenta proposta. A seguir, serão apresentados os resultados obtidos:

- "A ferramenta de análise automática de sites me ajuda a identificar facilmente as áreas do meu site que não estão em conformidade com a LGPD."
  - 93,3% (28) concordam totalmente;
  - 6,7% (2) concordam parcialmente;
  - 0% não concordam, nem discordam;
  - 0% discordam parcialmente;
  - 0% discordam totalmente.

Analisando as respostas com base nessa afirmação, pode-se inferir que de fato a ferramenta ajuda as pessoas a identificarem possíveis não conformidades com a LGPD em um *site*. Um número reduzido de participantes expressou concordância parcial com essa afirmação, observando que, embora a ferramenta identifique as não conformidades, pode haver casos em que seu desempenho não seja satisfatório para todos os usuários.

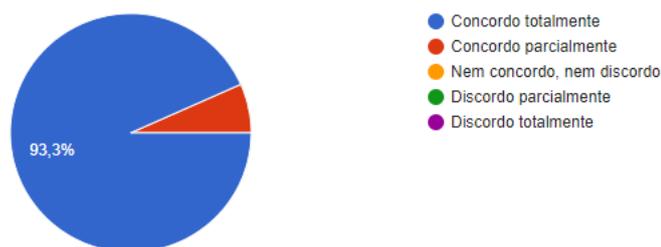


Figura 5.5: Gráfico das respostas sobre o auxílio da ferramenta na identificação de não conformidades com a LGPD. (Fonte: elaborada pelo autor).

- "Eu acredito que usar a ferramenta de análise automática me ajudará a melhorar a conformidade do meu site com a LGPD."
  - 93,3% (28) concordam totalmente;
  - 6,7% (2) concordam parcialmente;
  - 0% não concordam, nem discordam;
  - 0% discordam parcialmente;
  - 0% discordam totalmente.

Com base nos resultados obtidos nesta afirmação, é evidente que a ferramenta efetivamente contribui para melhorar a conformidade de um *site*, ao identificar não conformidades e sugerir modificações necessárias. Embora uma grande maioria concorde plenamente com a afirmação, uma minoria expressou concordância parcial. Como discutido na questão anterior, em certos casos, o desempenho da ferramenta pode não ser satisfatório e, portanto, não auxiliar o usuário na manutenção da conformidade de um *site*.

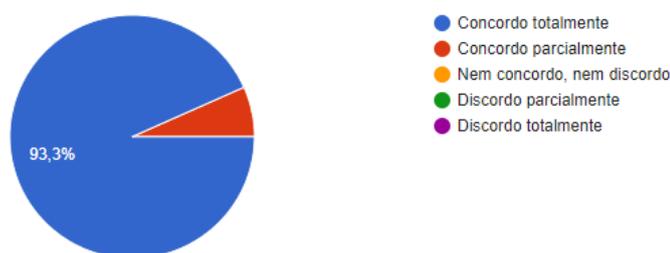


Figura 5.6: Gráfico das respostas sobre o auxílio da ferramenta na melhoria da conformidade de um site (Fonte: elaborada pelo autor).

- "A ferramenta de análise automática me permite compreender melhor algumas exigências da LGPD em relação ao meu site."
  - 80% (24) concordam totalmente;
  - 16,7% (5) concordam parcialmente;

- 3,3% (1) discordam parcialmente;
- 0% nem concordam, nem discordam;
- 0% discordam totalmente.

Analisando as respostas relacionadas a essa afirmação, observa-se que a maioria concorda que a ferramenta fornece informações necessárias e suficientes para entender melhor as exigências da lei e como segui-las. Entretanto, há ainda uma quantidade significativa que concorda ou discorda parcialmente dessa afirmação. Isso sugere que, apesar da existência de páginas dedicadas na ferramenta para explicar os conceitos da LGPD analisados e recomendar mudanças necessárias nos sites avaliados, alguns indivíduos consideram que essas informações não são totalmente abrangentes para compreender todas as exigências da lei.

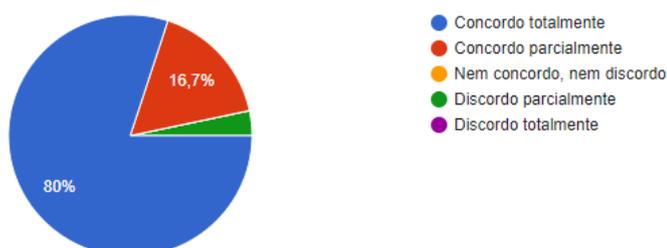


Figura 5.7: Gráfico das respostas sobre o auxílio da ferramenta na compreensão da LGPD (Fonte: elaborada pelo autor).

- "Usar a ferramenta de análise automática é benéfico para manter a conformidade contínua com a LGPD no meu site."
  - 90% (27) concordam totalmente;
  - 6,7% (2) concordam parcialmente;
  - 3,3% (1) nem concordam, nem discordam;
  - 0% discordam parcialmente;
  - 0% discordam totalmente.

Com base nas respostas a respeito dessa afirmação, fica claro que a maioria das pessoas concorda que a ferramenta proporciona benefícios na manutenção contínua da conformidade com a LGPD em um *site*. Todavia, uma minoria concorda parcialmente com essa afirmação, indicando que consideraram situações em que a ferramenta pode não ter um desempenho tão eficaz.

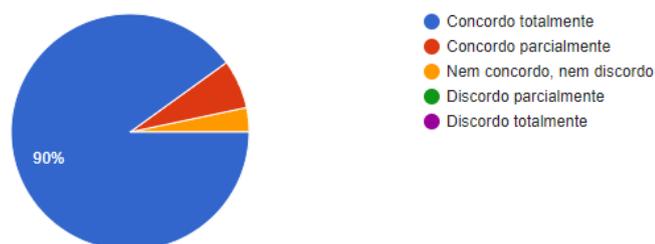


Figura 5.8: Gráfico das respostas sobre o benefício da ferramenta na manutenção da conformidade de um site (Fonte: elaborada pelo autor).

- "Usar a ferramenta de análise automática ajuda a evitar possíveis penalidades e multas por não conformidades com a LGPD."
  - 83,3% (25) concordam totalmente;
  - 13,3% (4) concordam parcialmente;
  - 3,3% (1) nem concordam, nem discordam;
  - 0% discordam parcialmente;
  - 0% discordam totalmente.

Analisando as respostas relacionadas a essa afirmação, conclui-se que a maioria dos participantes concorda que, ao auxiliar na identificação de não conformidades com a LGPD e sugerir modificações necessárias, a ferramenta pode ajudar o administrador do site a evitar possíveis penalidades por não estar em conformidade com a lei. No entanto, uma minoria expressou que concorda parcialmente com a afirmação, sugerindo que seria útil se a ferramenta também fornecesse informações sobre as multas e penalidades previstas pela lei.

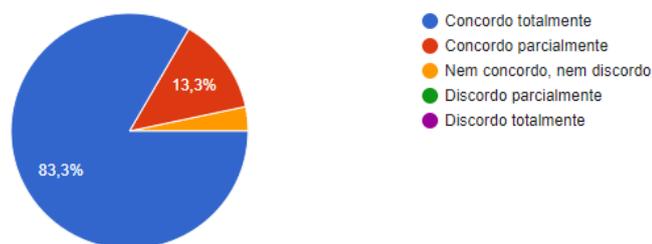


Figura 5.9: Gráfico das respostas sobre o auxílio da ferramenta em evitar multas por não conformidades em um site. (Fonte: elaborada pelo autor).

- "Se eu fosse um profissional do direito e precisasse identificar não conformidades em páginas de websites, a ferramenta de análise automática me ajudaria neste trabalho."
  - 90% (27) concordam totalmente;
  - 10% (3) concordam parcialmente;

- 0% nem concordam, nem discordam;
- 0% discordam parcialmente;
- 0% discordam totalmente.

É possível perceber, com base nas respostas sobre a afirmação, que a maioria concorda que a ferramenta realmente pode auxiliar um profissional da área jurídica na análise de um *site*. No entanto, uma minoria concorda parcialmente com essa afirmação, sugerindo que essas pessoas acreditam que, em alguns casos, o trabalho de análise manual por um profissional é superior à ferramenta e suficiente para a tarefa.

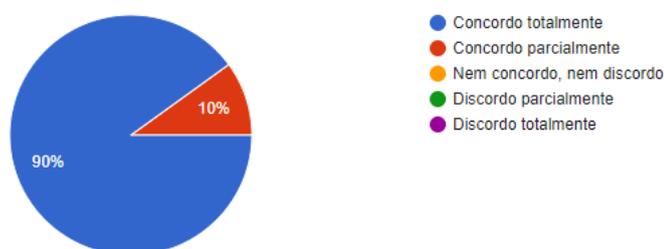


Figura 5.10: Gráfico das respostas sobre o auxílio da ferramenta no trabalho de um profissional de direito (Fonte: elaborada pelo autor).

- "Se eu fosse um profissional do direito e especialista em LGPD, eu utilizaria a ferramenta no meu trabalho."
  - 70% (21) concordam totalmente;
  - 23,3% (7) concordam parcialmente;
  - 6,7% (2) nem concordam, nem discordam;
  - 0% discordam parcialmente;
  - 0% discordam totalmente.

Analisando as respostas sobre essa afirmação, observa-se que uma grande parte das pessoas reconhece os benefícios que a ferramenta pode trazer para profissionais da área de LGPD e a importância de sua utilização nesse contexto. No entanto, uma parcela considerável respondeu que concorda parcialmente com a afirmação, sugerindo que em alguns casos elas percebem que a ferramenta pode não ser tão benéfica quanto esperado para esses profissionais. Além disso, uma minoria indicou não concordar nem discordar com a afirmação, possivelmente por considerarem não ter conhecimento suficiente para opinar sobre o assunto.

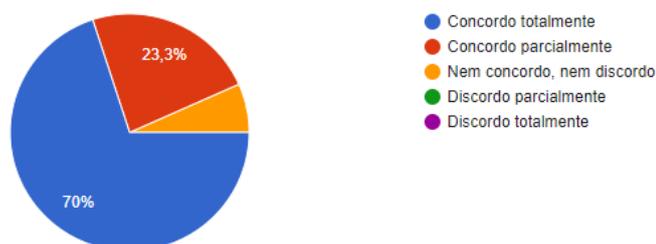


Figura 5.11: Gráfico das respostas sobre a adesão do uso da ferramenta no trabalho de um profissional de direito (Fonte: elaborada pelo autor).

- "A ferramenta de análise automática poderia facilmente substituir um especialista em direito e LGPD."
  - 46,7% (14) discordam totalmente;
  - 33,3% (10) nem concordam, nem discordam;
  - 16,7% (5) discordam parcialmente;
  - 3,3% (1) concordam parcialmente;
  - 0% concordam totalmente.

Com base nas respostas coletadas sobre essa afirmação, fica claro que ninguém acredita que a ferramenta possa substituir a análise de um especialista em LGPD atualmente. A maioria das pessoas manifestou discordância, indicando que a ferramenta ainda não atinge o nível de uma avaliação realizada manualmente por um profissional da área. Além disso, uma parte considerável dos participantes optou por não concordar nem discordar da afirmação, possivelmente por reconhecerem o potencial da ferramenta, mas não se sentirem suficientemente aptos para formarem uma opinião definitiva devido à falta de conhecimento.

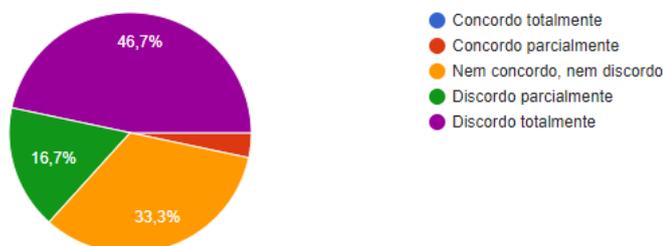


Figura 5.12: Gráfico das respostas sobre a substituição do especialista em direito pela ferramenta (Fonte: elaborada pelo autor).

- "A ferramenta de análise automática não seria um substituto para um especialista em direito e LGPD, mas serviria como uma segunda opinião, complementando a expertise humana na identificação de possíveis não conformidades com a legislação."

- 86,7% (26) concordam totalmente;
- 10% (3) concordam parcialmente;
- 3,3% (1) nem concordam, nem discordam;
- 0% discordam parcialmente;
- 0% discordam totalmente.

Com base nas respostas sobre essa afirmação, é perceptível que os participantes da pesquisa acreditam que, atualmente, a ferramenta serve efetivamente como um suporte ao profissional de LGPD, mas não tem a capacidade de substituí-lo.

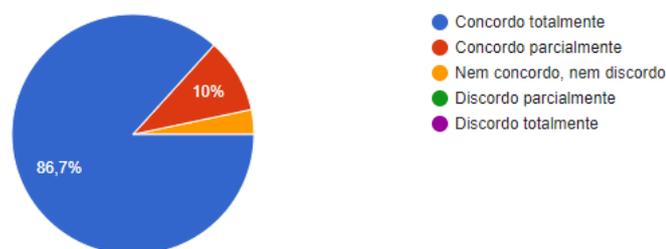


Figura 5.13: Gráfico das respostas sobre como a ferramenta não substitui o especialista, mas complementa seu trabalho (Fonte: elaborada pelo autor).

Portanto, ao analisar os resultados das percepções dos participantes, observa-se uma tendência à concordância com todas as afirmações positivas sobre a ferramenta, isso evidencia sua contribuição e relevância para o propósito para o qual foi desenvolvida, bem como aspectos da área jurídica.

Entretanto, nota-se a discordância com a afirmação relacionada à substituição do profissional especialista em LGPD pela ferramenta. Concorda-se com essa percepção, em decorrência da diversidade de cenários e da necessidade de interpretação humana. Isso indica que a ferramenta é percebida como um apoio aos especialistas em seus trabalhos, além de auxiliar administradores de *sites* com conhecimento limitado sobre a lei.

#### 5.4.2 Facilidade de uso da ferramenta

Para avaliar a opinião dos participantes sobre o nível de facilidade de uso da ferramenta no processo de análise, foram apresentadas afirmações, detalhadas com maior clareza no capítulo anterior (vide Capítulo 4). A seguir, apresentamos o resultado da percepção dos participantes:

- "É fácil aprender a usar a ferramenta de análise automática de sites para a LGPD."
  - 63,3% (19) concordam totalmente;

- 26,7% (8) concordam parcialmente;
- 6,7% (2) discordam parcialmente;
- 3,3% (1) nem concordam, nem discordam.
- 0% discordam totalmente.

Percebe-se que a maioria dos usuários acha relativamente fácil aprender a usar a ferramenta. Alguns mostraram uma concordância parcial com essa afirmação, sugerindo que conseguiram utilizá-la, mas encontraram algumas dificuldades. Uma pequena parte expressou discordância parcial, indicando que seria necessário fornecer instruções mais detalhadas sobre o uso da ferramenta.

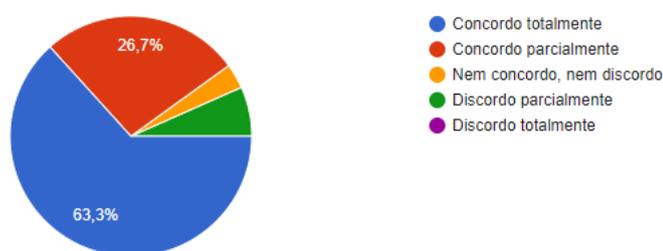


Figura 5.14: Gráfico das respostas sobre a facilidade em aprender a usar a ferramenta (Fonte: elaborada pelo autor).

- "A interface da ferramenta de análise automática é intuitiva e fácil de navegar."
  - 60% (18) concordam totalmente;
  - 33,3% (10) concordam parcialmente;
  - 6,7% (2) nem concordam, nem discordam;
  - 0% discordam parcialmente;
  - 0% discordam totalmente.

Observa-se que a maioria considera o uso da ferramenta fácil e intuitivo. Entre esses usuários, alguns expressaram concordância parcial com a afirmação, possivelmente porque acham a ferramenta fácil de usar, mas ainda têm algumas dúvidas sobre sua utilização. Além disso, alguns responderam de maneira neutra, possivelmente porque encontraram tanto aspectos positivos quanto negativos na interface.

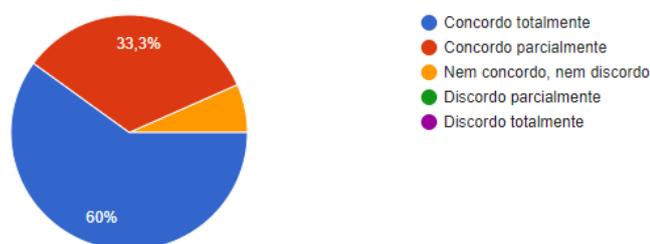


Figura 5.15: Gráfico das respostas sobre a facilidade de navegação na interface da ferramenta (Fonte: elaborada pelo autor).

- "A ferramenta fornece *feedback* imediato sobre os problemas de conformidade encontrados, o que facilita a correção."
  - 63,3% (19) concordam totalmente;
  - 33,3% (10) concordam parcialmente;
  - 3,3% (1) discordam parcialmente;
  - 0% nem concordam, nem discordam;
  - 0% discordam totalmente.

Nota-se que a maioria dos usuários acredita que a análise automatizada pela ferramenta é realizada de forma extremamente rápida, oferecendo um *feedback* quase imediato sobre a página analisada. No entanto, uma pequena parte dos que concordaram com essa afirmação indicaram uma concordância parcial, possivelmente devido a casos em que a ferramenta levou um pouco mais de tempo do que o habitual para concluir a análise. Além disso, uma minoria demonstrou discordância parcial, sugerindo que a ferramenta pode ter demorado excessivamente para realizar a análise ou que não forneceu *feedback* devido a problemas inesperados.

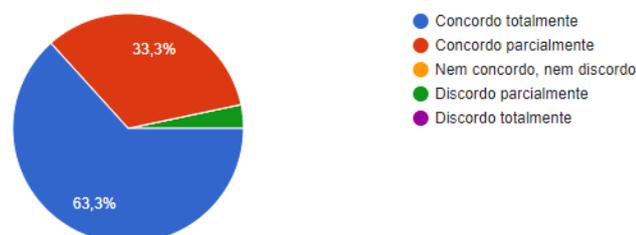


Figura 5.16: Gráfico das respostas sobre o *feedback* imediato da ferramenta (Fonte: elaborada pelo autor).

- "Eu posso usar a ferramenta facilmente sem a necessidade de treinamento extensivo ou suporte técnico."
  - 43,3% (13) concordam totalmente;

- 36,7% (11) concordam parcialmente;
- 13,3% (4) discordam parcialmente;
- 6,7% (2) nem concordam, nem discordam;
- 0% discordam totalmente.

Observa-se que a maioria dos usuários considera o uso da ferramenta relativamente simples, sugerindo que não é necessário um treinamento extensivo ou suporte técnico para sua operação. No entanto, alguns dos que concordaram parcialmente com essa afirmação indicaram que pessoas com menos familiaridade com tecnologias poderiam precisar de mais suporte. Além disso, uma quantidade significativa de usuários discordou parcialmente, apontando dificuldades no uso da ferramenta e sugerindo que um treinamento mais detalhado poderia ser benéfico.

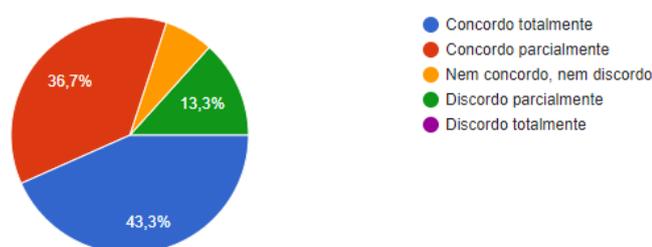


Figura 5.17: Gráfico das respostas sobre a necessidade de treinamento para o uso da ferramenta (Fonte: elaborada pelo autor).

- "O uso da ferramenta torna-se fácil e acessível a todos devido ao fato de ser uma aplicação *web*, não exigindo instalação nem configuração."
  - 83,3% (25) concordam totalmente;
  - 10% (3) nem concordam, nem discordam;
  - 6,7% (2) concordam parcialmente;
  - 0% discordam parcialmente;
  - 0% discordam totalmente.

A maioria dos voluntários concorda que, por ser uma aplicação *web*, a ferramenta é fácil e acessível de usar. No entanto, uma pequena parte dos participantes expressou concordância parcial com essa afirmação. Eles sugerem que, embora a maioria das pessoas tenha acesso à *internet* e, portanto, encontre a ferramenta facilmente, essa afirmação não se aplica para aqueles que têm acesso limitado ou inexistente à *internet*. Além disso, uma pequena parcela dos participantes indicou não ter conhecimento suficiente para formar uma opinião sobre a afirmação.

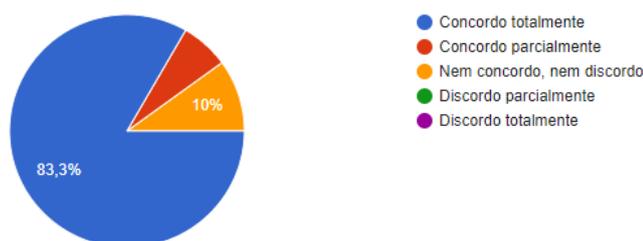


Figura 5.18: Gráfico das respostas sobre a acessibilidade da ferramenta (Fonte: elaborada pelo autor).

- "Eu não preciso ser um profissional de direito e especialista em LGPD para utilizar a ferramenta."
  - 80% (24) concordam totalmente;
  - 13,3% (4) concordam parcialmente;
  - 6,7% (2) nem concordam, nem discordam;
  - 0% discordam parcialmente;
  - 0% discordam totalmente.

A maioria das pessoas concordou que qualquer usuário pode utilizar a ferramenta e entender seus resultados sem precisar ser um especialista em LGPD. No entanto, alguns concordaram apenas parcialmente, sugerindo que, em certos casos, ainda seria necessário ter conhecimento especializado da lei para compreender totalmente os resultados. Uma pequena parte dos participantes preferiu adotar uma posição neutra, possivelmente porque acreditavam que apenas um especialista poderia avaliar adequadamente essa afirmação.

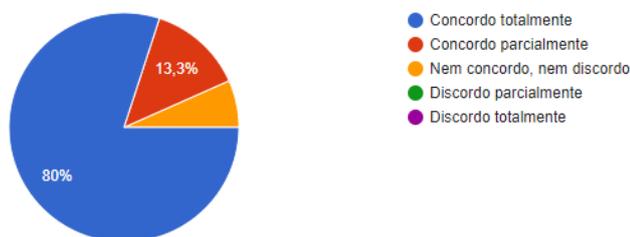


Figura 5.19: Gráfico das respostas sobre a necessidade de ser um profissional de direito para utilizar a ferramenta (Fonte: elaborada pelo autor).

- "Eu não preciso entender de programação e ciência da computação para utilizar a ferramenta."
  - 86,7% (26) concordam totalmente;
  - 6,7% (2) concordam parcialmente;

- 3,3% (1) nem concordam, nem discordam;
- 3,3% (1) discordam totalmente;
- 0% discordam parcialmente.

Percebe-se que a maioria dos participantes concordou que a utilização da ferramenta não requer conhecimentos prévios em programação ou ciência da computação. No entanto, uma pequena parcela dos entrevistados concordou parcialmente, enquanto outra discordou totalmente, e alguns mantiveram-se neutros. Não foi identificado um motivo claro para essas respostas divergentes, visto que a ferramenta é acessada por meio de uma interface gráfica em uma aplicação *web*, dispensando a necessidade de conhecimentos de ciência da computação ou de programação por parte do usuário.

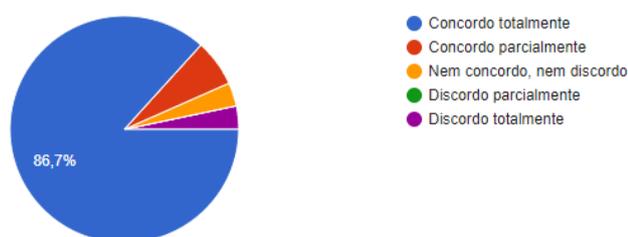


Figura 5.20: Gráfico das respostas sobre a necessidade do conhecimento em programação para utilizar a ferramenta (Fonte: elaborada pelo autor).

Com base nos resultados desta seção, observou-se que a maioria dos participantes concorda que aprender e utilizar a ferramenta de análise automatizada é uma tarefa fácil e acessível. Muitos destacaram a interface intuitiva e as funcionalidades claras da ferramenta como pontos positivos, que facilitam o processo de verificação de conformidade com a LGPD.

No entanto, é importante notar que, ainda que pequena, uma parcela de participantes expressou discordância ou uma opinião neutra a respeito deste tópico. Diante disso, considera-se que melhorias e desenvolvimento de recursos adicionais de suporte pode ser benéfico. Proporcionar um guia detalhado, como uma página de instruções ou um vídeo tutorial, pode ajudar a esclarecer procedimentos complexos e melhorar a experiência geral dos usuários ao utilizar a ferramenta. Essa medida não apenas promoveria uma maior adoção da tecnologia, mas também garantiria que todos os usuários possam maximizar o potencial da ferramenta.

## 5.5 Discussões sobre os resultados

Com base nos resultados discutidos neste capítulo, conclui-se que a ferramenta de análise automática de sites para verificação de conformidade com a LGPD efetivamente cumpre os objetivos propostos (conforme detalhado no Capítulo 3). Os dados mostram que a ferramenta é

significativamente mais rápida do que a análise manual realizada pelos participantes, conforme evidenciado pelo tempo médio de processamento. Além disso, a ferramenta demonstrou ser mais precisa do que a maioria das pessoas, com base na análise do erro médio na identificação de não conformidades pelos voluntários. Assim, a ferramenta poderá servir como um recurso valioso para especialistas em LGPD, oferecendo uma segunda opinião em seu trabalho, e também ajudará usuários comuns a entender melhor a lei e a identificar não conformidades em páginas ao redor da *web*.

No entanto, são necessárias melhorias para tornar a ferramenta mais eficaz em cenários reais. É fundamental desenvolver mecanismos mais sofisticados que ajudem usuários leigos a utilizar a ferramenta de forma adequada. Além disso, é essencial implementar novos *crawlers* que abordem cláusulas da lei não cobertas no escopo deste trabalho e novas regras mais específicas e complexas, com o objetivo de identificar o maior número possível de não conformidades em páginas de *sites*.

# 6

## Conclusão

O trabalho apresentado concentrou-se no desenvolvimento de uma ferramenta automatizada para avaliar a conformidade de *sites* com a Lei Geral de Proteção de Dados (LGPD) e incluiu a realização de um experimento com voluntários para coletar suas opiniões sobre a ferramenta.

A LGPD é uma legislação extremamente relevante atualmente (vide Capítulo 1) e cada vez mais, empresas e administradores de sites estão precisando se adaptar às suas cláusulas para evitar possíveis punições por não conformidade. Assim, foi preciso desenvolver uma ferramenta que realiza essas análises de forma rápida e eficaz, oferecendo suporte tanto para administradores de sites que precisam identificar não conformidades com a lei em suas páginas quanto para profissionais da área jurídica que realizam essas análises manualmente.

Assim, foi possível detalhar toda a construção da ferramenta, abrangendo desde a elaboração da arquitetura de *software* e a escolha das tecnologias utilizadas até a sua implementação e utilização na prática da ferramenta pelos voluntários que participaram do experimento. A coleta de opiniões dos voluntários foi fundamental para validar os objetivos propostos em termos de prova de conceito e para identificar possíveis áreas de aprimoramento e futuros trabalhos que possam enriquecer ainda mais esta pesquisa.

Assim, os resultados obtidos apontam que a ferramenta desenvolvida é significativamente mais rápida do que a análise realizada por seres humanos, demonstrando seu valor especialmente em situações que demandam análises em larga escala. Além disso, foi possível perceber que uma análise manual realizada por alguém sem especialização em LGPD tende a ser muito mais imprecisa em comparação com a análise automatizada fornecida pela ferramenta. Embora o experimento não tenha contado com a participação de um especialista na área, os voluntários expressaram a convicção de que a ferramenta será valiosa como suporte diário para os profissionais.

## 6.1 Trabalhos futuros

A seguir, será apresentada a lista de trabalhos de interesse que serão realizados no futuro:

- Construção de regras mais elaboradas nos *crawlers*, com o objetivo de trazer mais precisão à ferramenta;
- Desenvolvimento de novos *crawlers* para abranger o maior número possível de itens da LGPD;
- Elaboração de um plano detalhado e específico para suporte ao usuário, visando esclarecer todas as dúvidas sobre a utilização da ferramenta;
- Implementação de um modelo de inteligência artificial baseado em processamento de linguagem natural para identificar o contexto dos textos extraídos das páginas da *web* e analisar com maior precisão as não conformidades com a lei;
- Realização de estudo baseado em análise híbrida, na qual o profissional trabalha em conjunto com a ferramenta automatizada;
- Realização de um estudo focado na utilização da ferramenta exclusivamente por especialistas em direito e LGPD, com o objetivo de compreender melhor a perspectiva dos profissionais sobre a ferramenta.

# Referências bibliográficas

- [1] Redis. <https://redis.io/>, 2024. Acessado em 26 de fevereiro de 2024.
- [2] Luciane Cardoso Barzotto and Ricardo Hofmeister de Almeida Martins Costa. Estudos sobre lgpd—lei geral de proteção de dados—lei nº 13.709/2018: doutrina e aplicabilidade no âmbito laboral. 2022.
- [3] Brendon Benegas, Elenilson Lima de Freitas, Richard Alexandre Idesti Junior, Thalles Ariel de Oliveira, Evandro Jose Theodoro, and Marcia Regina Reggiolli. Proposta de adequação à lgpd: um estudo de caso hipotético. *Prospectus (ISSN: 2674-8576)*, 3(1):369–377, 2021.
- [4] Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. The world-wide web. *Communications of the ACM*, 37(8):76–82, 1994.
- [5] Brasil. Lei nº 13.709/2018 - lei geral de proteção de dados (lgpd), 2018. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm).
- [6] Felipe Timbó Brito and Javam Castro Machado. Preservação de privacidade de dados: Fundamentos, técnicas e aplicações. *Jornadas de atualização em informática*, pages 91–130, 2017.
- [7] Vergilio Ricardo Britto-da Silva, Edimara Mezzomo Luciano, and Odirlei Antonio Magnagnagno. Preocupação com a privacidade na internet: uma pesquisa exploratória no cenário brasileiro. *Anais do V Encontro de Administração da Informação, 2015, Brasil.*, 2015.
- [8] Aaron Cahn, Scott Alfeld, Paul Barford, and S. Muthukrishnan. An empirical study of web cookies. In *Proceedings of the 25th International Conference on World Wide Web, WWW '16*, page 891–901, Republic and Canton of Geneva, CHE, 2016. International World Wide Web Conferences Steering Committee. ISBN 9781450341431. DOI [10.1145/2872427.2882991](https://doi.org/10.1145/2872427.2882991). URL <https://doi.org/10.1145/2872427.2882991>.

- [9] Luiz Carvalho, Jonice Oliveira, Claudia Cappelli, and Violeta Majer. Desafios de transparência pela lei geral de proteção de dados pessoais. In *Anais do VII Workshop de Transparência em Sistemas*, pages 21–30. SBC, 2019.
- [10] Elizabeth Castro. *HTML for the world wide web*. Peachpit Press, 2003.
- [11] Sílvio César Cazella, MASN Nunes, and Eliseo Reategui. A ciência da opinião: Estado da arte em sistemas de recomendação. *André Ponce de Leon F. de Carvalho*, pages 161–216, 2010.
- [12] Anish Chapagain. *Hands-On Web Scraping with Python: Perform advanced scraping operations using various Python libraries and tools such as Selenium, Regex, and others*. Packt Publishing Ltd, 2019.
- [13] Luísa Margarida Antunes da Cunha et al. *Modelos Rasch e Escalas de Likert e Thurstone na medição de atitudes*. PhD thesis, 2007.
- [14] Escola Superior do Ministério Público da União. Guia rápido da lei geral de proteção de dados (lgpd). <https://escola.mpu.mp.br/transparencia/lei-geral-de-protECAo-de-dados/guiarapidolgpd.pdf>. Acessado em 15 de fevereiro de 2024.
- [15] Lara Rocha Garcia, Edson Aguilera-Fernandes, Rafael Augusto Moreno Gonçalves, and Marcos Ribeiro Pereira-Barretto. *Lei Geral de Proteção de Dados (LGPD): guia de implantação*. Editora Blucher, 2020.
- [16] Mohammad Taufiq Abdul Ghani, Mahizer Hamzah, Saipolbarin Ramli, Wan Ab, Aziz Wan Daud, Taj Rijal Muhamad Romli, and Nur Najihah Mohamad Mokhtar. A questionnaire-based approach on technology acceptance model for mobile digital game-based learning. *Journal of Global Business and Social Entrepreneurship (GBSE)*, 5 (14):11–21, 2019.
- [17] David Gourley and Brian Totty. *HTTP: the definitive guide*. "O'Reilly Media, Inc.", 2002.
- [18] Moaiad Ahmad Khder. Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3), 2021.
- [19] Robert C. Martin. *Clean Architecture: A Craftsman's Guide to Software Structure and Design*. Prentice Hall, 2017.
- [20] Ryan Mitchell. *Web scraping with Python: Collecting more data from the modern web*. "O'Reilly Media, Inc.", 2018.

- [21] Renato Leite Monteiro. Existe um direito à explicação na lei geral de proteção de dados do Brasil. *Artigo estratégico*, 39:1–14, 2018.
- [22] Sergio Nojiri. O direito à privacidade na era da informática: algumas considerações. *Revista Jurídica UNIJUS*, 8(8):99–106, 2005.
- [23] Emil Persson. Evaluating tools and techniques for web scraping, 2019.
- [24] Anderson Apolônio Lira Queiroz. A invasão de privacidade na internet: um modelo de boas práticas e uma proposta interativa de proteção da privacidade por meio dos cookies. Master's thesis, Universidade Federal de Pernambuco, 2011.
- [25] Cláudio Filipe Lima Rapôso, Haniel Melo de Lima, Waldecy Ferreira de Oliveira Junior, Paola Aragão Ferreira Silva, and Elaine Elaine de Souza Barros. Lgpd-lei geral de proteção de dados pessoais em tecnologia da informação: Revisão sistemática. *RACE-Revista de Administração do Cesmac*, 4:58–67, 2019.
- [26] Gabriel Luis Andriago Pio Romão and Daisy Eboli. Lgpd descomplicado. *REVISTA FOCO*, 15(7):e617–e617, 2022.
- [27] Isabela Maria Rosal Santos. Lgpd: Quais são os direitos do titular dos dados? 2021.
- [28] Valdemar W Setzer. Dado, informação, conhecimento e competência. *DataGramZero Revista de Ciência da Informação*, n. 0, 28, 1999.
- [29] Lipi Shah, Hetal Gaudani, and Prem Balani. Survey on recommendation system. *International Journal of Computer Applications*, 137(7):43–49, 2016.
- [30] Maurício Samy Silva. *Criando sites com HTML: sites de alta qualidade com HTML e CSS*. Novatec Editora, 2008.
- [31] De S Sirisuriya et al. A comparative study on web scraping. 2015.
- [32] Raphaela Velho. Em vigor a partir de agosto, implementação da lei geral de proteção de dados ainda enfrenta desafios. *Ciência e Cultura*, 72(2):09–11, 2020.
- [33] Yabo Xu, Ke Wang, Benyu Zhang, and Zheng Chen. Privacy-enhancing personalized web search. In *Proceedings of the 16th international conference on World Wide Web*, pages 591–600, 2007.