

**UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO PROFISSIONAL EM GESTÃO DA INFORMAÇÃO E DO
CONHECIMENTO**

OCANTE ANTÓNIO IÉ

**APARATO DOS ALGORITMOS NA CIÊNCIA DA INFORMAÇÃO:
COMPREENDENDO OS PROCESSOS DE AUTOMAÇÃO DOS MOTORES DE BUSCA**

**SÃO CRISTÓVÃO-SE
2023**

OCANTE ANTÔNIO IÉ

**APARATO DOS ALGORITMOS NA CIÊNCIA DA INFORMAÇÃO:
COMPREENDENDO OS PROCESSOS DE AUTOMAÇÃO DOS MOTORES DE BUSCA**

Dissertação apresentada à Universidade Federal de Sergipe, como parte das exigências do Programa de Pós-Graduação em Ciência da Informação como pré-requisito para obtenção do título de mestre em Gestão da Informação e do Conhecimento.

Orientadora: Profa. Dra. Alessandra dos Santos Araújo.

Co-Orientadora: Profa. Dra. Martha Suzana Cabral Nunes

**SÃO CRISTÓVÃO-SE
2023**

Dados de Catalogação na Publicação (CIP)

I22a	Ié, Ocante António. Aparato dos algoritmos na Ciência da Informação [manuscrito] : compreendendo os processos de automação dos motores de busca / Ocante António Ié. – São Cristóvão, 2023. 107f. il. color. Orientadora: Profa. Dra. Alessandra dos Santos Araújo. Co-orientadora: Profa. Dra. Martha Suzana Cabral Nunes. Dissertação (Mestrado Profissional em Gestão da Informação e do Conhecimento) – Universidade Federal de Sergipe, Programa de Pós-graduação em Ciência da Informação, 2023. 1. Ciência da Informação. 2. Algoritmos. 3. Recuperação da informação. I. Araújo, Alessandra dos Santos, orientadora. II. Nunes, Martha Suzana Cabral, co-orientadora. III. Título. CDU 02-004.021:519.683.5 CDD 003
------	--

Ficha elaborada pela bibliotecária Joyce Dayse de Oliveira Santos (CRB-5/SE-002005).

OCANTE ANTÔNIO IÉ

**APARATO DOS ALGORITMOS NA CIÊNCIA DA INFORMAÇÃO:
COMPREENDENDO OS PROCESSOS DE AUTOMAÇÃO DOS MOTORES DE BUSCA**

Dissertação apresentada à Universidade Federal de Sergipe, como parte das exigências do Programa de Pós-Graduação em Ciência da Informação como pré-requisito para obtenção do título de mestre em Gestão da Informação e do Conhecimento.

Avaliação: APROVADO

Data da defesa: 31/10/2023

BANCA EXAMINADORA

Documento assinado digitalmente



ALESSANDRA DOS SANTOS ARAÚJO
Data: 13/12/2023 19:01:01-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Alessandra dos Santos Araújo
(Orientadora)

Profa. Dra. Martha Suzana Cabral Nunes
(Co-Orientadora)

Documento assinado digitalmente



ALESSANDRA CABRAL NOGUEIRA LIMA
Data: 14/12/2023 08:57:27-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Alessandra Cabral Nogueira Lima
(Membro externo)

Documento assinado digitalmente



PABLO BOAVENTURA SALES PAIXÃO
Data: 13/12/2023 19:22:16-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Pablo Boaventura Sales Paixão
(Membro interno)

Dedico esta obra à minha família, uma fonte inesgotável de amor e inspiração. Às vezes, o coração se parte, mas é desse quebrar que construímos algo verdadeiramente extraordinário. A minha mãe, meu pilar de amor e coragem, que me apoiou desde o início, mesmo à distância. Às minhas irmãs e irmãos, cujos sorrisos e palavras de encorajamento foram meu combustível diário. Lembro com carinho do meu pai, que partiu, mas cujo amor e sabedoria ainda me guiam. E ao meu sobrinho, que partiu cedo demais, mas que sempre será um lembrete do valor da vida e da resiliência. Apesar das perdas, dificuldades e distâncias, esta dedicação reflete a jornada em busca de conhecimento e melhores condições para nossa família. Esta dissertação é uma expressão do amor e da certeza de que podemos fazer algo grandioso, contribuindo para a sociedade que tanto valorizamos. Que seja uma homenagem a ciência, à perseverança e à determinação de todos que acreditaram em mim, mostrando que, com dedicação e motivação, podemos alcançar nossos objetivos, não importa quão distantes.

Àqueles que me inspiraram e tornaram esta jornada possível, dedico este trabalho com profundo respeito e gratidão. Que estas páginas sejam um tributo ao amor e ao apoio que moldaram o meu caminho.

AGRADECIMENTOS

A realização deste trabalho representa não apenas um marco em minha jornada acadêmica, mas também um momento de profunda gratidão que desejo compartilhar com todos que estiveram ao meu lado nesta trajetória de aprendizado e descoberta.

Primeiramente, quero expressar minha mais sincera gratidão aos professores do Programa de Pós-graduação em Ciência da Informação e Gestão da Informação e do Conhecimento da Universidade Federal de Sergipe. Suas mentes brilhantes, orientação e paixão pelo conhecimento foram fundamentais para a minha formação. Cada aula, discussão e conselho que recebi moldaram meu pensamento e me inspiraram a alcançar novos patamares na minha pesquisa e estudos.

Aos meus orientadores, e a toda a equipe que me apoiou, agradeço por sua orientação dedicada e paciência. Suas orientações foram o farol que guiou meu caminho ao longo deste mestrado. Suas valiosas sugestões e conselhos moldaram este trabalho de pesquisa, e sou profundamente grato por sua disponibilidade constante em me ajudar a superar desafios e alcançar meus objetivos acadêmicos.

Aos meus queridos colegas de curso, que compartilharam comigo o entusiasmo pela ciência e formaram uma comunidade de aprendizado, agradeço pelas discussões enriquecedoras, colaborações e amizades que construímos. Cada um de vocês contribuiu de maneira única para minha experiência acadêmica e pessoal.

Além disso, quero estender meu agradecimento a todos os que, de maneira direta ou indireta, contribuíram para este trabalho e minha jornada no mestrado. Seja através de conversas enriquecedoras, recursos compartilhados, ou simplesmente estando presente em momentos de desafio, sua influência positiva não passou despercebida.

Neste momento de realização, estou ciente de que essa conquista é o resultado de um esforço coletivo e do apoio generoso de tantas pessoas. O conhecimento adquirido ao longo deste mestrado não apenas enriqueceu minha mente, mas também expandiu minha compreensão sobre a importância da colaboração e do apoio mútuo.

Este trabalho de mestrado é, portanto, dedicado a todos vocês, com profundo apreço e gratidão. Que nossas jornadas continuem a se cruzar, e que possamos continuar a contribuir para o crescimento do conhecimento e da sociedade.

"Os algoritmos são o esqueleto subjacente de todos os nossos avanços em inteligência artificial, aprendizado de máquina e ciência de dados. Eles são os segredos que estão impulsionando a revolução da informação."

Pedro Domingos (2015).

RESUMO

A pesquisa teve como objetivo compreender o comportamento dos algoritmos nos processos de automação dos motores de busca na perspectiva da Ciência da Informação, utilizando a abordagem metodológica da *Design Science Research*. Buscou-se explorar as funcionalidades dos algoritmos e suas implicações na sociedade contemporânea, visando identificar novas ideias e aprimorar o acesso à informação. A interação entre a Ciência da Informação e os algoritmos de busca é fundamental para otimizar a recuperação da informação, compreender os fluxos informacionais e possibilitar a produção de sentido no contexto digital. A pesquisa justificou-se pela necessidade de explorar os avanços tecnológicos e as estratégias de disponibilização de dados, bem como a importância da Ciência da Informação em adaptar-se às transformações no campo da informação. A metodologia adotada foi a *Design Science*, que busca criar soluções práticas para os desafios identificados, contribuindo para o desenvolvimento do campo da Ciência da Informação e para a otimização dos processos de busca de informações na era digital. O Web Crawler desenvolvido como produto resultante desta pesquisa representa uma solução tangível para o desafio de compreender os algoritmos de busca na web na perspectiva da Ciência da Informação (CI). Sua aplicação prática e o impacto potencial que detém na pesquisa em CI enfatizam sua significativa relevância e contribuição para o progresso do conhecimento nesta área. Através desta ferramenta, é possível explorar as funcionalidades dos algoritmos e suas implicações, promovendo uma melhor compreensão dos fluxos informacionais na era digital e facilitando o acesso à informação de maneira mais eficaz e precisa.

Palavras-chave: automação de motores de busca; algoritmos; Ciência da Informação; acesso à informação.

ABSTRACT

The research aims to understand the behavior of algorithms in search engine automation processes from the perspective of Information Science, using the Design Science methodological approach. The aim is to explore the functionalities of algorithms and their implications in contemporary society, aiming to identify new ideas and improve access to information. The interaction between Information Science and search algorithms is fundamental to optimize information retrieval, understand information flows and enable the production of meaning in the digital context. The research is justified by the need to explore technological advances and data availability strategies, as well as the importance of Information Science in adapting to transformations in the field of information. The methodology adopted is Design Science, which seeks to create practical solutions to the identified challenges, contributing to the development of the field of Information Science and the optimization of information search processes in the digital age. The Web Crawler developed as the resultant product of this research represents a tangible solution to the challenge of understanding web search algorithms from the perspective of Information Science (IS). Its practical application and the potential impact it holds in IS research emphasize its significant relevance and contribution to the advancement of knowledge in this field. Through this tool, it is possible to explore the functionalities of algorithms and their implications, promoting a better understanding of information flows in the digital age and facilitating access to information in a more effective and precise manner.

Keywords: search engine automation; algorithms; Information Science, access to information.

LISTA DE FIGURAS

Figura 1	- Fluxograma Estrutural dos Algoritmos.....	48
Figura 2	- Comportamento de Motores de Buscas.....	53
Figura 3	- Estrutura de Arquivo.....	86
Figura 4	- Ativação de <i>Environment</i>	87
Figura 5	- Banco de Dados.....	88
Figura 6	- Extração de Informação Via Console.....	89
Figura 7	- <i>Template</i> da aplicação – Página de pesquisa (<i>Data Search</i>)	90
Figura 8	- Resultados das buscas via navegador - 1.....	91
Figura 9	- Resultados das buscas via navegador - 2.....	91
Figura 10	- Resultados das buscas via navegador - 3.....	92
Figura 11	- Título Resultados das buscas via navegador - 4.....	92
Figura 12	- Resultados Filtradas e Armazenadas.....	93
Figura 13	- Classificação e Organização de Informações no Banco de Dados.....	94

LISTA DE QUADROS

Quadro 1	- Conceitos e Descrições da Metodologia DSR.....	60
Quadro 2	- Diferenças entre a Ciência Tradicional e a <i>Design Science</i>	62
Quadro 3	- Os Componentes, Pesquisa Tradicional e DSR.....	63
Quadro 4	- Aplicação da Diretriz.....	65
Quadro 5	- Elemento da Pesquisa.....	75
Quadro 6	- Principais pontos na análise SWOT.....	82

LISTA DE ABREVIATURAS E SIGLAS

ADI	<i>American Documentation Institute</i>
ASIS	<i>American Society for Information Science</i>
ASIST	<i>American Society for Information Science and Technology</i>
CI	Ciência da Informação
CIKM	<i>Conference on Information and Knowledge Management</i>
ERIC	<i>Educational Resources Information Center</i>
FID	Federação Internacional de Documentação
GB	<i>Google Busca</i>
GIC	Gestão da Informação e do Conhecimento
GS	<i>Google Search</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IA	Inteligência Artificial
IBBD	Instituto Brasileiro de Bibliografia e Documentação
Ibict	Instituto Brasileiro de Informação em Ciência e Tecnologia
IBM	<i>International Business Machines</i>
IIB	Instituto Internacional de Bibliografia
IID	Instituto Internacional de Documentação
IoT	<i>Internet das Coisas (Internet of Things)</i>
IUPAC	<i>International Union on Pure and Applied Chemistry</i>
KDD	<i>Knowledge Discovery in Databases</i>
OSDI	<i>Ocular Surface Disease Index</i>
PPGCI	Programa de Pós-graduação em Ciência da Informação
RI	Recuperação de Informações
SERP	Resultados do Mecanismo de Pesquisa (<i>Search Engine Results Page</i>)
SIGIR	<i>Special Interest Group on Information Retrieval Information Retrieval</i>
SIGMOD	<i>Special Interest Group on Management of Data</i>
SMART	Sistema para Análise Mecânica e Recuperação de Texto
SO	Sistema Operacional
SWOT	<i>Strengths, Weaknesses, Opportunities and Threats</i>
TIC	Tecnologias da Informação e da Comunicação
UFS	Universidade Federal de Sergipe

URL	<i>Uniform Resource Locator</i>
VLDB	<i>Very Large Database</i>
WSDM	<i>Web Search and Data Mining</i>
WWW	<i>World Wide Web</i>

SUMÁRIO

1	INTRODUÇÃO.....	15
1.1	Objetivos.....	16
1.1.	Objetivo geral.....	
1		16
1.1.	Objetivos específicos.....	
2		16
1.2	Justificativa.....	17
2	FUNDAMENTAÇÃO TEÓRICA.....	19
2.1	Ciência da Informação, evolução, definição e campo de atuação.....	20
2.2	Documentação: Relação com Ciência da Informação e Algoritmos.....	29
	.	
2.3	Relações Interdisciplinares envolvendo a Ciência da Informação e o estudo dos Algoritmos.....	33
3	ALGORITMOS DE BUSCA NA WEB.....	37
3.1	Algoritmos e suas Estruturas.....	45
3.2	Funcionamento de Motores de Buscas.....	49
3.3	A Busca Automatizada Utilizando a Linguagem <i>Python</i>.....	57
3.4	<i>Design Science Research</i>.....	59
3.5	Aplicação de Diretrizes de <i>Design Science Research</i>.....	64
3.5.	A Importância do Problema.....	
1		65
3.5.	Avaliação de Artefatos.....	
2		66
3.5.	Rigor da Pesquisa.....	
3		67
3.5.	Avaliação da Pesquisa.....	
4		69
3.5.	Comunicação da Pesquisa.....	
5		71
4	METODOLOGIA.....	75

4.1	Tipo de Pesquisa e Abordagem.....	75
4.2	Método da Pesquisa.....	76
4.3	Técnica de Coleta de Dados.....	77
4.4	Técnica de Tratamento de Dados.....	78
4.5	Teste como Técnica de Tratamento de Dados.....	78
4.6	Análise SWOT.....	79
4.6.	Pontos Fortes (<i>Strengths</i>).....	
1		80
4.6.	Pontos Fracos (<i>Weaknesses</i>).....	
2		81
4.6.	Oportunidades (<i>Opportunities</i>).....	
3		81
4.6.	Ameaças (<i>Threats</i>).....	
4		81
5	PRODUTO.....	83
5.1	Documentação.....	
.		84
5.2	Descrição da Estrutura dos Arquivos.....	85
5.3	Execução da Aplicação.....	86
5.4	Banco de Dados	87
5.5	Execução do Arquivo <i>App.Py</i> para iniciar a Aplicação <i>Flask</i>	88
5.6	Etapas do Processo de Teste para o Projeto.....	89
5.7	Aplicação <i>web</i> de <i>web</i> <i>scraping</i>	
.		90
5.8	Teste de Integração Entre Componentes.....	92
5.9	Análise de Dados e Resultados	94
6	CONSIDERAÇÕES FINAIS.....	97
	REFERÊNCIAS.....	10
		1

1 INTRODUÇÃO

Sabe-se que os avanços tecnológicos no âmbito da comunicação e informação exigem cada vez mais novas estratégias de disponibilização de dados e difusão da informação, buscando automatizar sempre os mecanismos de busca na *web*. Em meio a essa realidade, encontra-se a tecnologia digital, considerada como uma das maiores tendências inovadoras da revolução contemporânea.

No cenário contemporâneo, a interseção entre tecnologia, algoritmos e ciência da informação desempenha um papel crucial na evolução da sociedade. Segundo Tocc et al. (2007), os avanços tecnológicos têm influência direta na qualidade das informações e na forma como interagimos com os documentos digitais.

A rápida ascensão da inteligência artificial e o desenvolvimento de algoritmos avançados, como ressaltado por Harari (2019) em suas obras mais recentes, estão redefinindo não apenas a maneira como processamos informações, mas também moldando as estruturas sociais e econômicas. A habilidade de analisar grandes conjuntos de dados e extrair insights significativos têm potencial transformador, impulsionando avanços em campos que vão desde a medicina até a economia. Sobre essa vertente, essa tecnologia tem impregnado o modo de interação com os documentos, proporcionando novos meios e recursos para a criação cultural e para a dinamização das relações sociais. No âmbito da ciência da informação, o autor destaca como a capacidade de gerenciar e interpretar dados torna-se cada vez mais vital.

Esses avanços tecnológicos estão presentes no cotidiano, nas ciências, nas tecnologias e, principalmente, na maioria das informações acessadas nas plataformas virtuais. De acordo com Araújo (2018), a convivência com a documentação na *web* de dados é inevitável, dada sua ampla presença e seu impacto na sociedade. Nesse contexto, a Ciência da Informação (CI) deve se adequar e aderir a esse meio, compreendendo as propriedades da informação e os processos de sua construção, comunicação e uso.

O problema de pesquisa que orienta este estudo está relacionado à compreensão do comportamento dos algoritmos nos processos de automação dos motores de busca na perspectiva da CI. Um algoritmo pode ser definido como uma sequência finita de passos lógicos e bem definidos para a resolução de um problema ou execução de uma tarefa específica. Segundo O'Neil (2016), algoritmos são conjuntos de regras ou instruções para realizar uma tarefa ou resolver um problema, muitas vezes implementados por computadores para automatizar processos e tomar decisões. A crescente influência dos avanços tecnológicos na qualidade das informações e na forma como interagimos com documentos digitais levanta

questões fundamentais sobre como esses algoritmos funcionam e como eles afetam a organização, recuperação e apresentação de informações. Portanto, o problema de pesquisa pode ser formulado da seguinte maneira:

Como os algoritmos utilizados nos motores de busca automatizados impactam a seleção, classificação e apresentação de informações na era digital, e de que maneira a CI pode compreender e responder a esses desafios, contribuindo para o avanço do campo científico e auxiliando na tomada de decisões na sociedade contemporânea?

Este problema de pesquisa direciona a investigação para a compreensão dos algoritmos de busca e sua influência na disponibilização de informações, estabelecendo uma base sólida para os objetivos específicos delineados no estudo.

1.1 Objetivos

Ao alcançar esses objetivos, espera-se obter um conhecimento aprofundado sobre o comportamento dos algoritmos nos processos de automação dos motores de busca, contribuindo para a construção de um arcabouço teórico-prático na área da CI e fornecendo compreensão relevante para o campo acadêmico e profissional.

1.1.1 Objetivo geral

Diante dessa realidade, esta pesquisa teve como objetivo geral compreender o comportamento dos algoritmos nos processos de automação dos motores de busca na perspectiva da CI, utilizando a *Design Science* como abordagem metodológica, visando contribuir para o avanço do campo científico e auxiliar a tomada de decisões na sociedade contemporânea.

1.1.2 Objetivos específicos

- Investigar as estruturas dos algoritmos utilizados nos motores de busca, analisando sua composição e funcionamento, com foco na compreensão das técnicas e processos que orientam a indexação e recuperação de informações;
- Analisar de maneira abrangente o funcionamento dos mecanismos de busca empregados na web, explorando os fatores e sinais considerados pelos algoritmos para classificar e apresentar os resultados de pesquisa. Esta análise se concentrará em identificar as variáveis que influenciam a relevância e a ordenação dos resultados;
- Desenvolver e implementar um exemplo de *Web Crawler* utilizando a linguagem de programação *Python*, com o propósito de realizar buscas automatizadas de informações na web, a fim de compreender a interação entre o processo de automação e os algoritmos de busca.

Os objetivos específicos deste estudo visam aprofundar a compreensão da interseção entre algoritmos de busca e automação, destacando sua relevância na Ciência da Informação (CI). A investigação detalhada das estruturas e funcionamentos dos algoritmos utilizados nos motores de busca proporciona insights valiosos para pesquisadores interessados na evolução das tecnologias de busca e seu impacto na organização e acesso à informação. A análise

abrangente dos mecanismos de busca na web, considerando os fatores que influenciam a relevância e a ordenação dos resultados, contribui para o entendimento crítico dessas tecnologias na Ciência da Informação.

Além disso, a implementação prática do Web Crawler, como proposto no terceiro objetivo, oferece uma análise concreta da aplicabilidade dessas tecnologias. Destacando sua relevância, o produto resultante contribui diretamente para as práticas de pesquisa em Ciência da Informação, promovendo uma compreensão mais aplicada e abrangente das inovações tecnológicas. Ao permitir buscas automatizadas de informações na web, o estudo proporciona uma visão tangível de como a automação e os algoritmos de busca podem ser utilizados, enriquecendo o campo da CI com perspectivas práticas e aplicáveis. Esta abordagem integrada entre teoria e prática busca fortalecer a relação entre a evolução tecnológica e as necessidades informacionais, ampliando o escopo de pesquisa na área de Ciência da Informação.

Os objetivos específicos deste estudo visam aprofundar a compreensão da interseção entre algoritmos de busca e automação, destacando sua relevância na Ciência da Informação (CI). A investigação detalhada das estruturas e funcionamentos dos algoritmos utilizados nos motores de busca proporciona insights valiosos para pesquisadores interessados na evolução das tecnologias de busca e seu impacto na organização e acesso à informação. A análise abrangente dos mecanismos de busca na web, considerando os fatores que influenciam a relevância e a ordenação dos resultados, contribui para o entendimento crítico dessas tecnologias na Ciência da Informação.

Além disso, a implementação prática do Web Crawler, como proposto no terceiro objetivo, oferece uma análise concreta da aplicabilidade dessas tecnologias. Destacando sua relevância, o produto resultante contribui diretamente para as práticas de pesquisa em Ciência da Informação, promovendo uma compreensão mais aplicada e abrangente das inovações tecnológicas. Ao permitir buscas automatizadas de informações na web, o estudo proporciona uma visão tangível de como a automação e os algoritmos de busca podem ser utilizados, enriquecendo o campo da CI com perspectivas práticas e aplicáveis. Esta abordagem integrada entre teoria e prática busca fortalecer a relação entre a evolução tecnológica e as necessidades informacionais, ampliando o escopo de pesquisa na área de Ciência da Informação.

Adicionalmente, esses objetivos específicos estão alinhados com as demandas atuais da pesquisa em CI, que busca compreender os desafios e oportunidades apresentados pela crescente automatização dos processos de busca de informação. A investigação das estruturas e do funcionamento dos algoritmos dos motores de busca é fundamental para compreender como a informação é selecionada, classificada e apresentada aos usuários,

proporcionando compreensão sobre o papel dos algoritmos na organização e recuperação da informação.

1.2 Justificativa

A interação entre a CI e os algoritmos de busca é de fundamental importância para otimizar a recuperação da informação, compreender os fluxos informacionais e possibilitar a produção de sentido no contexto digital. Compreender esses aspectos contribui para a melhoria dos sistemas de busca e para a eficácia na gestão da informação na sociedade contemporânea. Além disso, essa pesquisa se justifica pela necessidade de explorar os avanços tecnológicos e as novas estratégias de disponibilização de dados, bem como a importância da CI em se adaptar e compreender as transformações no campo da informação.

Assim, esta pesquisa buscou avançar no conhecimento sobre os algoritmos dos processos de automação dos motores de busca, considerando sua relevância na sociedade atual e a necessidade de aprimoramento dos sistemas de busca e Gestão da Informação. Por meio da abordagem do *Design Science Research* como metodologia de pesquisa, espera-se criar soluções práticas para os desafios identificados, contribuindo para o desenvolvimento do campo da CI e para a otimização dos processos de busca de informações na era digital.

A abordagem de *Design Science Research* (DSR) refere-se a uma metodologia de pesquisa que concentra seus esforços na criação de soluções práticas e inovadoras para problemas complexos do mundo real. Ao invés de apenas investigar e descrever fenômenos, como em muitos métodos tradicionais de pesquisa, o DSR visa projetar, implementar e avaliar artefatos tangíveis que abordam esses problemas. Esses artefatos podem ser sistemas, modelos, métodos ou processos.

Na pesquisa mencionada sobre os algoritmos dos processos de automação dos motores de busca, a abordagem de Design Science foi empregada para não apenas compreender a fundo esses algoritmos, mas também para criar soluções práticas e aplicáveis. Em vez de apenas teorizar sobre melhorias nos sistemas de busca e Gestão da Informação, a pesquisa buscou desenvolver e implementar soluções tangíveis, contribuindo assim para avançar o conhecimento nesse campo e otimizar os processos de busca de informações na era digital. A abordagem de Design Science Research visa, portanto, traduzir o conhecimento teórico em ações concretas e impactantes no mundo real.

A Justificativa desta pesquisa será apresentada para destacar sua relevância e seu alinhamento com as demandas contemporâneas em CI, enfatizando a importância crítica da

compreensão dos algoritmos de busca na era digital. Na seção sobre *Design Science Research*, explicaremos a abordagem metodológica escolhida, justificando-a e delineando os princípios e etapas do processo de pesquisa. Seguindo de Metodologia, onde detalhará os métodos que foram utilizados para coleta e análise de dados, descrevendo a abordagem para investigar as estruturas dos algoritmos e o funcionamento dos motores de busca, bem como o desenvolvimento do *Web Crawler* em *Python*. O *Web Crawler* é uma ferramenta automatizada que percorre a web de forma sistemática, coletando informações relevantes. No contexto deste estudo, o *Web Crawler* em *Python* foi elaborado como um exemplo prático para compreender a interação entre automação e algoritmos de busca, permitindo uma análise concreta da aplicabilidade dessas tecnologias. Essa ferramenta não apenas ilustrou a teoria, mas também contribuiu para o aprimoramento das práticas de pesquisa na área da Ciência da Informação.

Em seguida, aplicou-se a análise *Strengths, Weaknesses, Opportunities and Threats* (SWOT) à pesquisa, identificando os pontos fortes, pontos fracos, oportunidades e ameaças do estudo.

Seguindo com apresentação do Produto, será exposto o exemplo de *Web Crawler* desenvolvido, demonstrando sua funcionalidade e explicando como ele contribui para a compreensão dos algoritmos de busca, além de mostrar sua relevância para os objetivos da pesquisa. Na Conclusão, faremos uma reflexão abrangente sobre os principais resultados e contribuições da pesquisa, discutindo implicações teóricas e práticas, contribuindo para pesquisas futuras na área de CI e automação de busca. Essa estrutura permite discutir as implicações mais amplas dos achados, sua aplicação prática em diversos domínios e destacando eventuais limitações e direções possíveis para a pesquisa. Diante disso, a abordagem garante uma organização completa e coesa do estudo, abrangendo todos os aspectos necessários para uma análise aprofundada do comportamento dos algoritmos nos motores de busca.

A dissertação está estruturada em uma sequência cuidadosamente organizada para abordar de maneira completa e coesa o tema da pesquisa. Se inicia com a seção de Fundamentação Teórica, aprofundaremos os conceitos fundamentais em Ciência da Informação e nos algoritmos de busca, conduzindo uma revisão detalhada da literatura e explorando as tendências e avanços nessa área. Na seção sobre *Design Science Research*, explicamos a abordagem metodológica escolhida, justificando-a e delineando os princípios e etapas do processo de pesquisa. A seção de Metodologia detalhou os métodos que foram utilizados para coleta e análise de dados, descrevendo a abordagem para investigar as estruturas dos algoritmos e o funcionamento dos motores de busca, bem como o desenvolvimento do *Web Crawler* em

Python. Em seguida, aplicamos a análise SWOT à pesquisa, identificando os pontos fortes, pontos fracos, oportunidades e ameaças do estudo.

Em seguida com apresentação do Produto, foi exposto o exemplo de *Web Crawler* desenvolvido, demonstrando sua funcionalidade e explicando como ele contribuiu para a compreensão dos algoritmos de busca, além de mostrar sua relevância para os objetivos da pesquisa. Finalmente, na seção de Conclusão, foi realizada uma reflexão abrangente sobre os principais resultados e contribuições da pesquisa, discutindo implicações teóricas e práticas e sugerindo direções para pesquisas futuras na área de Ciência da Informação e automação de busca. Essa estrutura foi concluída com a seção de Considerações Finais, que desempenha um papel fundamental no encerramento do estudo, permitindo-nos discutir as implicações mais amplas dos achados, sua aplicação prática em diversos domínios e destacando eventuais limitações e direções futuras para a pesquisa. Essa abordagem garante uma organização completa e coesa do estudo, abrangendo todos os aspectos necessários para uma análise aprofundada do comportamento dos algoritmos nos motores de busca.

2 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica deste estudo buscou-se estabelecer uma base sólida de conhecimento no campo da CI e dos algoritmos dos motores de busca sem redundâncias. Por meio de uma revisão da literatura, exploramos os conceitos-chave, as discussões relevantes que sustentam a compreensão do comportamento dos algoritmos e sua aplicação na recuperação da informação.

No primeiro tópico, abordamos os principais conceitos relacionados à CI, incluindo sua evolução, definição e âmbito de atuação. Examinamos as perspectivas teóricas sobre o ato de informar, a construção e comunicação da informação, bem como a importância da gestão e organização da informação na sociedade contemporânea. Realizamos uma análise aprofundada dos algoritmos dos motores de busca, destacando suas características, funcionalidades e relevância na recuperação da informação e na otimização dos processos de busca. Exploramos as abordagens teóricas e os estudos existentes sobre esses algoritmos, ressaltando suas implicações na sociedade atual e os desafios relacionados à eficácia na busca de informações.

No contexto da CI, a compreensão dos processos de automação dos motores de busca assume um papel fundamental. Portanto, a fundamentação teórica abordou os princípios e conceitos relacionados aos algoritmos de rastreamento, indexação e classificação utilizados por esses motores. Investigamos as técnicas e metodologias empregadas nesses processos de coleta de dados na *web*. Além disso, examinamos as teorias que fundamentam a recuperação da informação e a relação dos algoritmos de busca com esses modelos, como o modelo *Booleano*, o modelo *Vetorial* e o modelo *Probabilístico*.

Na sequência, exploramos as discussões sobre a transparência e a responsabilidade dos algoritmos de busca, abordando preocupações éticas relacionadas à classificação e ordenação dos resultados de busca, bem como a importância de garantir a diversidade de informações e evitar vieses algorítmicos. Dessa forma, a fundamentação teórica deste estudo proporcionará uma compreensão aprofundada do comportamento dos algoritmos dos motores de busca na perspectiva da CI e sua influência na recuperação da informação na sociedade contemporânea.

Adicionalmente, é relevante mencionar que, segundo Toscani e Veloso (2001), o termo "algoritmo" pode ser compreendido como uma sequência de raciocínios, instruções ou operações com a finalidade de alcançar um objetivo, desde que esses passos sejam finitos e sistematicamente executados. Menezes (2001) traz outra perspectiva, argumentando que um algoritmo se justifica pelo resultado que busca alcançar, o que implica ter um objetivo específico. Conforme Menezes (2001), a complexidade de um algoritmo pode crescer à medida que a necessidade de considerar diferentes cenários se apresenta, tornando-o mais intrincado para abranger todas as possibilidades. No entanto, assim como em códigos, é necessário percorrer o algoritmo passo a passo para atingir o objetivo final.

2.1 Ciência da Informação, evolução, definição e campo de atuação

Com o advento da ciência moderna no século XVI, surgiram as sociedades científicas, onde os pesquisadores se reuniam para compartilhar seus estudos e descobertas por meio de correspondências, buscando preservar o direito autoral de suas experiências. Segundo Pinheiro (2002), essas trocas representaram os primeiros indícios dos periódicos científicos, que consolidaram o processo de comunicação e informação no meio acadêmico no final do século XVII, estabelecendo as bases para a disseminação formal do conhecimento. Foi em 1665, com a criação do *Journal des Sçavans* na França e o *Philosophical Transactions* na Inglaterra, que se estabeleceu de forma oficial o processo de comunicação e informação científica, antecipando os periódicos científicos contemporâneos. Dessa forma, Meadows, (1999), afirma que o aumento do número de pesquisadores e, conseqüentemente, da produção científica, combinado com o progresso tecnológico, desempenhou um papel determinante no surgimento da CI.

A CI teve seu surgimento marcado por uma série de eventos e marcos históricos, alguns desses marcos incluem a criação do Instituto Internacional de Bibliografia (IIB) em 1895 durante a I Conferência Internacional de Bibliografia, liderada por Paul Otlet e Henri de La Fontaine. O IIB, conforme observado por Pinheiro (2002), tinha como objetivo estabelecer uma biblioteca universal do conhecimento, buscando democratizar o acesso à informação. Posteriormente, em 1931, o IIB foi transformado no Instituto Internacional de Documentação (IID) durante a X Conferência Internacional de Bibliografia, ainda sob a influência de Otlet e La Fontaine.

Outro marco importante foi a publicação da obra "*Traité de Documentation: le Livre sur le Livre: Théorie et Pratique*" por Paul Otlet em 1935, onde foram introduzidas ideias

sobre a integração dos componentes da documentação, além de abordar conceitos relacionados à Bibliometria. Em 1937, foi fundado o *American Documentation Institute* (ADI), que posteriormente se tornou a *American Society for Information Science* (ASIS) e, nos anos 2000, a *American Society for Information Science and Technology* (ASIST). Além disso, o IID foi transformado em Federação Internacional de Documentação (FID) em 1938, sendo palco de pesquisas teóricas que fundamentaram a CI.

De acordo com Pinheiro (2002), a CI emergiu em resposta a diversos fatores. Primeiramente, a "explosão da informação" decorrente do avanço científico e tecnológico demandado pela Segunda Guerra Mundial e o papel dos periódicos científicos tiveram um papel fundamental. Além disso, a necessidade de registro e transmissão de informações e conhecimentos, assim como o surgimento de novas tecnologias, como o computador, contribuíram para o desenvolvimento da CI.

Eventos como a Conferência de Informação Científica da *Royal Society* em 1948, a Conferência do *International Union on Pure and Applied Chemistry* (IUPAC) em 1955 e a Conferência Internacional de Informação Científica da Academia Nacional de Ciências em 1958 também desempenharam um papel significativo no surgimento da CI. Além disso, trabalhos acadêmicos, como o artigo sobre trabalho de informação de *Farradane* e o glossário de termos da documentação científica de Taylor (1951), bem como avanços tecnológicos, como o uso do microfilme e dos cartões perfurados da *International Business Machines* (IBM), contribuíram para a consolidação dessa nova disciplina. A criação de periódicos especializados, como o *Journal Documentation* em 1945, o *Nachrichten für Dokumentation* em 1950 e o *American Documentation* em 1950, também desempenharam um papel fundamental no desenvolvimento e disseminação da CI.

São considerados como marcos para o advento da CI, conforme Pinheiro (2002):

- I) Criação do IIB, em 1895, durante a I Conferência Internacional de Bibliografia, em Bruxelas. O IIB nasceu a partir das ideias de Paul Otlet (considerado o "pai" da CI) e Henri de La Fontaine, que desejavam organizar o "livro universal do conhecimento" (Mattelart, 2005). A intenção era a criação de uma biblioteca universal, com os registros bibliográficos de todos os documentos indexados do mundo, como se fosse uma biblioteca de referência e não de acervo (Oliveira, 2005), desse modo, democratizando a informação;
- II) Transformação do IIB em IID, em 1931, por sugestão de Paul Otlet e Henri de La Fontaine, durante a X Conferência Internacional de Bibliografia, em Bruxelas;

- III) Publicação da obra *Traité de Documentation: le Livre sur le Livre: Théorie et Pratique*, de Paul Otlet, em 1935, que trouxe ideias de integração dos componentes da documentação, já apresentando ideias a respeito da Bibliometria;
- IV) Fundação do ADI, em 1937 (mais tarde, virou ASIS; nos anos 2000 transformou-se em ASIST);
- V) Transformação do IID em FID, em 1938. Nesta Federação, desenvolveram-se pesquisas teóricas que foram as bases científicas da CI. Ainda de acordo com Pinheiro (2002), a CI nasceu:
 - Da "explosão da informação", decorrente do avanço científico e tecnológico demandado pela Segunda Guerra Mundial, bem como dos periódicos científicos (Russo, 2010);
 - Da necessidade de registro (controle bibliográfico) e transmissão de informação e conhecimento (serviços informacionais dispostos para atender, de acordo com Miranda (2002), à pesquisa e desenvolvimento);
 - Do surgimento de novas tecnologias, sobretudo o computador (que passou a ser empregado no processamento de informações bibliográficas já na década de sessenta).

Contudo, Barreto (2007), observa que os computadores começaram a ter o seu uso na CI somente no início dos anos oitenta. Diante dessa expansão informacional, a preocupação pelo registro e transmissão de informação e conhecimento foi descrita em influente artigo por Vannevar Bush (criador do Memex), "*As we may think*", de 1945. Neste artigo, importante na gênese da CI, o autor expressa a importância da preservação e armazenamento de documentos notáveis para a ciência, a fim de que fossem disponibilizados para consulta. Alguns eventos também contribuíram para o surgimento da CI, a se destacar: Conferência de Informação Científica, da *Royal Society*, em 1948, na cidade de Londres; Conferência do IUPAC, em Londres, no ano de 1955; e Conferência Internacional de Informação Científica, da Academia Nacional de Ciências, em Washington, no ano de 1958. Alguns trabalhos que, da mesma forma, contribuíram para o nascimento da CI: O artigo a respeito do trabalho de informação, de Farradane, e glossário de termos da documentação científica, de Taylor (1951).

Do mesmo modo, novas tecnologias colaboraram para o advento da CI, tais como: o microfilme (seu uso expandiu o espaço físico das bibliotecas, levando à criação do ADI) e os cartões perfurados, da IBM. Ainda colaboraram para o aparecimento da CI - os periódicos:

Journal Documentation, em 1945, na Grã-Bretanha; *Nachrichten für Dokumentation*, em 1950, na antiga URSS; e *American Documentation*, em 1950, nos Estados Unidos (Pinheiro, 2002).

A CI recebeu influência de diversas teorias e obras que contribuíram para o seu desenvolvimento e consolidação como campo de estudo. Entre essas influências, destacam-se as teorias de Norbert Wiener, Claude Shannon, Warren Weaver e Ludwig von Bertalanffy.

Norbert Wiener, por meio de sua Teoria da Informação e de sua obra "*Cybernetics or control and communication in the animal and the machine*" publicada em 1947, trouxe importantes contribuições para a CI. Sua teoria abordava a comunicação e o controle em sistemas, estabelecendo fundamentos para o entendimento da informação como um elemento essencial na interação entre seres humanos e máquinas.

Claude Shannon e Warren Weaver, por sua vez, influenciaram a CI com a Teoria Matemática da Comunicação ou Teoria da Informação, publicada em 1948. Essa teoria introduziu conceitos fundamentais, como ruído, entropia e redundância, que são essenciais para os sistemas de recuperação de informação. No entanto, é importante ressaltar que essa teoria foi originalmente desenvolvida para máquinas e não considerava completamente as características humanas de seleção e interpretação da informação.

Ludwig von Bertalanffy contribuiu com a CI por meio de sua Teoria Geral dos Sistemas, publicada em 1956. Essa teoria introduziu conceitos relacionados a redes e sistemas de informação, ampliando a compreensão dos processos informacionais como parte de sistemas mais amplos. Essa abordagem sistêmica foi fundamental para o desenvolvimento da CI como um campo multidisciplinar que busca compreender os diversos aspectos relacionados à informação e ao seu uso.

Na década de 1960 a CI emerge como campo de estudo, impulsionada principalmente pela recuperação da informação. Saracevic (1996) destaca que o surgimento formal da CI ocorreu nos Estados Unidos, entre os anos de 1961 e 1962. Nesse período, ocorreram avanços significativos, como a transição dos cartões perfurados para o uso de CD-ROM e acesso *on-line*, a evolução de sistemas não interativos para sistemas interativos, a mudança de bases documentais para bases de conhecimento e a transição de textos escritos para *multimídia*. Esses avanços refletiram uma ampliação das possibilidades de recuperação da informação e de interação com os sistemas informacionais.

Portanto, é possível afirmar que o surgimento da CI está intimamente ligado ao estudo dos problemas relacionados à recuperação da informação, bem como às contribuições teóricas de Wiener, Shannon, Weaver e Bertalanffy, que forneceram bases sólidas para o desenvolvimento desse campo de estudo.

De acordo com Miranda (2002), o surgimento da CI está intimamente relacionado à necessidade de lidar com a produção científica, sua disseminação, uso e obsolescência. Essa visão ressalta a importância do controle e da compreensão dos fenômenos relacionados à informação científica, incluindo sua dispersão, propagação e os diferentes aspectos envolvidos em sua manipulação e análise. Essa abordagem destaca a importância da CI na busca por estratégias e métodos que possibilitem o acesso, a organização e a utilização eficiente da informação em diferentes contextos.

O surgimento da CI está intrinsecamente relacionado à necessidade de lidar com a recuperação da informação diante do crescimento exponencial e da diversidade de informações surgidas a partir da década de 1940. Esse campo de estudo envolve diversas disciplinas, como Biblioteconomia, Informática, Psicologia e Comunicação, que contribuem para a compreensão dos processos de recuperação da informação. A pesquisadora Lena Vania Ribeiro Pinheiro identifica três fases na evolução da CI: a primeira, de 1962 a 1969, caracterizada por discussões iniciais sobre a origem, denominação, conceitos e definições da área; a segunda, de 1970 a 1989, marcada pela busca por princípios, metodologia e teorias próprias, considerando as transformações trazidas pelas novas tecnologias; e a terceira, a partir de 1990, que compreende a consolidação da área, a discussão de sua natureza e as relações interdisciplinares. No Brasil, a CI foi introduzida na década de 1970, com a criação do curso de Mestrado em CI pelo Instituto Brasileiro de Bibliografia e Documentação (IBBD), atualmente conhecido como Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict).

O periódico CI, criado em 1972, também contribuiu para a consolidação dessa área de estudo. No início da CI, havia confusão entre Biblioteconomia e Informática, o que levou a diferentes denominações, como Informação Científica, Ciência da Biblioteca e de Informação, Ciência e Tecnologia da Informação, Ciência e Engenharia da Informação, entre outras. Antes do surgimento da CI, os termos Bibliografia e Documentação eram amplamente utilizados, sendo a Bibliografia empregado nas bibliotecas e documentação nos centros de documentação que atendiam empresas, indústrias, pesquisadores e especialistas (Dias, 2002; Pinheiro, 2002; Russo, 2010).

A CI é uma disciplina que surgiu a partir das ideias de Taylor (1951), discutidas em reuniões realizadas no *Georgia Institute of Technology*, nos Estados Unidos, entre 1961 e 1962. Com base nessas ideias, foi desenvolvido um conceito semelhante, o de Taylor (1951), que se tornou amplamente aceito na literatura. Nesse sentido, CI investiga as propriedades e o comportamento da informação, as forças que governam seu fluxo e os significados do processamento da informação, com o objetivo de torná-la acessível e utilizável de forma ideal.

A disciplina abrange o conhecimento relacionado à origem, coleta, organização, armazenamento, recuperação, interpretação, transmissão, transformação e uso da informação. Isso inclui a análise da representação da informação em sistemas naturais e artificiais, o uso de códigos para a transmissão eficiente da mensagem e o estudo do processamento de informações e técnicas aplicadas a computadores e sistemas de programação. O conceito formulado no final da década de 1960, já evidencia a preocupação da CI com o acesso e a usabilidade da informação, questões que continuam sendo importantes até hoje e que são objeto de estudo nessa área. O conceito também destaca a natureza Interdisciplinar da CI, que se fundamenta em teorias da Biblioteconomia, Comunicação, Informática, Psicologia e outras disciplinas (Barreto, 2007).

A CI é uma disciplina que surgiu a partir das ideias de Taylor (1951), discutidas em reuniões realizadas no *Georgia Institute of Technology*, nos Estados Unidos, em 1961 e 1962, e investiga as propriedades e o comportamento da informação, as forças que governam seu fluxo e os significados do processamento da informação, com o objetivo de torná-la acessível e utilizável de forma ideal.

A CI, ao longo de sua história, tem sido abordada por diferentes pesquisadores, cada um trazendo contribuições importantes para a compreensão dessa área. Segundo Mikhailov e Giljarevskij (1970, p. 14), a CI é uma disciplina científica que investiga as estruturas, propriedades e regularidades da informação científica, abrangendo também suas teorias, história, metodologia e organização.

A ideia do caráter científico é enfatizada por esses autores, como evidenciado pelo uso frequente do termo "científico" em seu conceito. Destacando a importância de teorias e padrões na CI, como as leis bibliométricas, cientometria e informetria. Pinheiro (1997) identifica princípios fundamentais, como a obsolescência/vida média da literatura, análise de citações e a lei do elitismo.

Le Coadic (2004) destaca elementos que conferem à CI um caráter científico e social, incluindo revistas científicas, bancos de dados, recursos da *internet*, sociedades científicas e profissionais, bem como cursos e unidades de ensino na área.

Shera e Cleveland (1995 *apud* Braga, 1995, p. 4) definem a CI como uma ciência que investiga as propriedades, comportamento e fluxo da informação, além dos meios de processamento para garantir sua acessibilidade e usabilidade ótimas. Essa definição destaca a interdisciplinaridade, uma ideia que se diferencia do enfoque que enfatiza mais a transmissão da informação.

Saracevic (1996, p. 47), renomado pesquisador nessa área, descreve a CI como um campo dedicado à comunicação efetiva do conhecimento entre as pessoas, considerando o contexto social, institucional ou individual e as necessidades de informação. Ele ressalta o interesse nas vantagens das modernas tecnologias informacionais no tratamento dessas questões.

De acordo com Almeida, Bastos e Bittencourt (2007, p. 72), a CI é um campo social de produção do conhecimento voltado para questões relacionadas à informação. No entanto, é importante ressaltar que o conhecimento é resultado do processo de apropriação da informação pelo sujeito, conforme apreende seu sentido e o associa ao seu conhecimento prévio.

Em síntese, a CI é uma disciplina científica que investiga as estruturas, propriedades e regularidades da informação científica. Ela está fundamentada em teorias e padrões, abrangendo diferentes áreas do conhecimento. Além disso, busca a comunicação efetiva do conhecimento e está atenta às vantagens das tecnologias informacionais modernas.

A CI é uma área que abrange tanto o campo da informação quanto o da comunicação. Silva (2006) destaca a ideia de “infocomunicação” em sua definição, afirmando que é uma ciência social que investiga os problemas e fenômenos relacionados ao fluxo, organização e comportamento da informação. Isso engloba desde a geração, coleta, organização, armazenamento, recuperação, interpretação, transformação até a utilização da informação.

Outros estudiosos da área, como Capurro e Hjørland (2007), também enfatizam a importância dessas atividades na CI. Eles afirmam que a disciplina lida com a geração, coleta, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação, com ênfase na aplicação de tecnologias modernas nesses processos. Além disso, a CI busca criar e estruturar um corpo de conhecimento científico e tecnológico relacionado à transferência de informação.

É interessante notar que esses conceitos enfatizam a importância das tecnologias na área da CI. A utilização de tecnologias modernas é considerada fundamental para lidar com os desafios relacionados à informação. A CI é um campo em constante evolução, e novas definições e abordagens continuam a surgir.

Uma definição mais recente e abrangente da CI é apresentada por Bawden e Robinson (2012), destaca que a CI é o estudo interdisciplinar da geração, coleta, organização, armazenamento, recuperação, disseminação, uso e impacto da informação em suas diversas formas, tanto no contexto individual quanto social. Isso inclui a investigação das propriedades

e comportamentos da informação, bem como as teorias, métodos e práticas que sustentam sua criação, gestão e utilização efetivas.

Essa definição ressalta a natureza interdisciplinar da CI, além disso, destaca a importância de compreender não apenas as características da informação em si, mas também o seu impacto na sociedade e a forma como é utilizada.

Outra definição relevante é proposta por Tuominen, Savolainen e Talja (2018), que destacam a natureza social da informação e a necessidade de uma abordagem interdisciplinar. Segundo eles, a CI é uma disciplina que estuda a informação como um fenômeno social, investigando tanto o uso e o significado da informação nas diferentes áreas do conhecimento, como também os sistemas, práticas e políticas relacionadas à criação, organização, disseminação e uso da informação.

É interessante observar que, apesar das diferentes abordagens e nuances nas definições propostas, há pontos em comum entre elas. A maioria das definições reconhece a CI como uma disciplina e/ou área científica voltada para o estudo da informação em suas diversas dimensões, incluindo a geração, organização, disseminação, acesso, uso e impacto da informação na sociedade. Além disso, muitas definições destacam a importância das Tecnologias da Informação e da Comunicação (TIC) no contexto da CI.

Ela abrange tanto as propriedades e comportamentos da informação como também as teorias, métodos e práticas relacionadas à sua criação, gestão e utilização efetivas. Reconhece-se a importância da dimensão social da informação e a necessidade de uma abordagem interdisciplinar para compreender as práticas informacionais e seu contexto sociocultural.

O objeto da CI tem passado por mudanças ao longo do tempo. Segundo Dias (2002), inicialmente o foco estava na "orientação para o usuário", mas posteriormente evoluiu para o "uso da informação" e, mais tarde, para a "satisfação das necessidades individuais de informação". Atualmente, o objeto mais adequado, de acordo com o autor, é o "acesso à informação", que engloba o acesso financeiro, físico, intelectual e social à informação.

Le Coadic (2004) argumenta que o objeto da CI não é mais a biblioteca, o livro, o centro de documentação ou o objeto físico, mas sim a própria informação. A informação pode ser encontrada em documentos impressos, em diálogos entre cientistas, em patentes, em fotografias, em objetos, entre outros meios (Pinheiro, 2002). Le Coadic (2004) também destaca que a CI estuda as propriedades gerais da informação, como sua natureza, gênese e efeitos, bem como analisa os processos de construção, comunicação e uso da informação.

A comunicação desempenha um papel fundamental na troca de informações entre as pessoas, sendo o processo intermediário que permite essa troca (Le Coadic, 2004). Segundo Tavares (1998) a comunicação é o suporte que garante a informação, e ainda afirma que a informação é um fator de crescimento, progresso e produtividade, e os meios de comunicação desempenham o papel de estimular os processos cognitivos individuais, despertando a criatividade, a percepção e a ousadia.

Percebe-se a indissociabilidade entre informação e comunicação, sendo a informação o objeto da CI e a comunicação o meio para a existência da informação. Dessa forma, a CI estuda tanto o fluxo da comunicação como seus atores e os registros que transportam a informação e o conhecimento. Nesse sentido, o objetivo da CI é compreender a mecânica processual da comunicação, as instituições envolvidas, os produtos gerados, os especialistas, as ferramentas e as técnicas utilizadas, a fim de garantir a satisfação das necessidades de produção, transformação, utilização e comunicação do conhecimento.

A CI possui um foco abrangente, envolvendo tanto abordagens sociológicas quanto epistemológicas, voltadas para a geração, coleta, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação. (Capurro; Hjørland, 2007). Os objetivos estão relacionados ao tratamento da informação, incluindo a descrição e classificação da informação presente em diversos tipos de materiais, como livros, periódicos, filmes e mapas (Dias, 2002).

Goffman (1970) destaca que o objetivo da CI é estabelecer um enfoque científico para o estudo dos fenômenos relacionados à informação, presentes em processos biológicos, existência humana e máquinas criadas por humanos. Nesse sentido, a disciplina busca estabelecer princípios fundamentais que orientem o comportamento nos processos de comunicação e seus sistemas de informação associados.

Wersig e Neveling (1975) afirmam que a CI se baseia nas necessidades informacionais das pessoas envolvidas no trabalho social e estuda os métodos de organização dos processos de comunicação para atender a essas necessidades.

A CI é caracterizada pela sua natureza interdisciplinar, colaborando com várias áreas do conhecimento. De acordo com Le Coadic (2004), ela se relaciona com disciplinas como direito, economia, eletrônica, estatística, filosofia, informática, linguística, lógica, matemática, política, psicologia, sociologia e telecomunicações. Saracevic (1996) destaca sua relação mais direta com a biblioteconomia, a ciência da computação, a ciência cognitiva e a comunicação.

Dessa forma, a CI apresenta características gerais, conforme as ideias de Saracevic (1996): interdisciplinaridade em constante evolução, utilização da tecnologia da informação e contribuição para a evolução da sociedade da informação. A interdisciplinaridade decorre da variedade de profissionais que se envolveram com a CI desde seus primeiros tempos, contribuindo para sua evolução sob diferentes perspectivas (Saracevic, 1996).

Ao longo de sua evolução, a CI tem buscado compreender os fluxos informacionais e os mecanismos de busca como parte integrante do campo de estudo. A disciplina vem explorando os aspectos teóricos e práticos relacionados à recuperação da informação, desde a definição de modelos e métodos até a análise dos impactos sociais e éticos desses processos. Nesse sentido, a pesquisa sobre os algoritmos dos motores de busca se insere nesse contexto, contribuindo para o avanço da CI ao oferecer uma visão mais aprofundada sobre o comportamento desses algoritmos e suas implicações na sociedade contemporânea.

A compreensão dos algoritmos dos motores de busca na perspectiva da CI amplia as possibilidades de atuação nesses contextos, permitindo o desenvolvimento de estratégias mais eficientes de recuperação da informação, o aprimoramento dos serviços oferecidos e a facilitação do acesso à informação relevante. Portanto, o estudo desses algoritmos é de extrema relevância para o campo da CI, contribuindo para a sua evolução e fornecendo subsídios teóricos e práticos para a gestão e organização da informação na sociedade contemporânea.

No campo da CI, entender os processos de automação dos motores de busca é de fundamental importância. Como mencionado anteriormente, os motores de busca utilizam algoritmos de rastreamento, indexação e *ranking* para organizar e apresentar os resultados de pesquisa aos usuários. Esses algoritmos são baseados em princípios e conceitos específicos que permitem aos motores de busca analisar e classificar milhões de páginas da *web* em questão de segundos.

2.2 Documentação: Relação com Ciência da Informação e Algoritmos

Ao direcionar o foco para os objetos presentes em mídias e ambientes digitais dentro do contexto da CI, é relevante considerar as reflexões de Ferrez (1994), que já apontava naquela época que tais veículos de informação fornecem a base para se tornarem fontes de pesquisa científica e comunicação, gerando e disseminando informações inovadoras. Essa perspectiva ressalta a necessidade de um sistema de documentação que permita a organização ágil e precisa das informações, por meio da utilização de técnicas anteriormente inexistentes. No âmbito da pesquisa que visa compreender o comportamento dos algoritmos nos processos

de automação dos motores de busca, a aplicação desses conceitos da CI se torna ainda mais relevante, uma vez que os algoritmos são responsáveis por estruturar e fornecer acesso às informações presentes nas plataformas digitais. Portanto, a compreensão desses processos algorítmicos e a análise da documentação existente são fundamentais para uma investigação científica abrangente e embasada.

No contexto da pesquisa que busca compreender o comportamento dos algoritmos dos processos de automação dos motores de busca na perspectiva da CI, é importante destacar a relevância da documentação como disciplina fundamental nesse campo. De acordo com os estudos de Rayward (1997), Silva e Freire (2012) e Albagli (2015), a própria formação da CI está intrinsecamente ligada à documentação. Silva e Freire (2012) ressaltam que a necessidade de organização, registro e difusão da informação e seus fluxos é o fundamento primordial para o surgimento dessa área de estudo.

Ao abordar a documentação, é válido mencionar que ela esteve originalmente associada à Biblioteconomia do século XV até o final do XIX, quando passou a ser desenvolvida por Paul Otlet e La Fontaine como uma articulação entre os sistemas práticos de informação e a estruturação teórica (Rayward, 1997). A publicação do "*Traité de Documentation*" em 1934 por Otlet tornou-se um marco importante, em que o termo "Documentação" passou a ser associado ao novo domínio do conhecimento, abrangendo princípios documentais relacionados a diversos aspectos como registro, conservação, circulação, organização, indexação, resumos e fichamento (Ortega; Lara, 2010; Silva; Freire, 2012).

Essa relação entre a Documentação e a CI resalta a importância de compreender a evolução e os fundamentos teóricos dessas áreas para uma análise mais abrangente do comportamento dos algoritmos dos processos de automação dos motores de busca. A documentação fornece as bases para a estruturação e organização eficiente da informação, sendo um elemento essencial para compreender a maneira como os motores de busca processam e disponibilizam os dados aos usuários. A pesquisa nesse campo permite uma análise mais aprofundada e embasada dessas tecnologias e sua aplicação na CI.

As reflexões de Otlet e La Fontaine (2010) acerca da ampliação da noção de suportes de informação, reconhecendo praticamente tudo como um documento que representa uma ação humana ou um detalhe da natureza, estabelecem um diálogo relevante com o tema da pesquisa em questão. Essa perspectiva, fundamentada em subsídios institucionais, teórico-práticos e teórico-bibliográficos, busca focalizar o fenômeno social da informação e sua conexão com a CI.

À medida que ocorrem mudanças na conceituação da documentação *web* e no objeto de estudo das *mídias* digitais, o entendimento do objeto se expande, compreendendo-o como um repositório de múltiplos níveis informacionais complexos (Meadows, 1999). Nesse contexto, a *web* de dados direciona sua análise não apenas ao objeto físico e às coleções, mas ao conteúdo informacional presente nesses objetos. Decarli e Tzagaraki (2006) enfatizam a importância dos processos que qualificam os dados na *web* como ferramentas para pesquisa, preservação e comunicação. Quando a informação fornecida pela documentação é devidamente organizada e tratada para promover sua acessibilidade, ela adquire a capacidade de atender a diferentes públicos e diversos fins.

Esse pensamento converge com a temática central da CI, como mencionado por Pinheiro (2005, p.40). A CI reconhece a necessidade da documentação e organização da informação para torná-la útil e acessível a uma ampla gama de usuários. Dessa forma, ao compreender o comportamento dos algoritmos nos processos de automação dos motores de busca na perspectiva da CI, é fundamental considerar essas abordagens teóricas e práticas relacionadas à documentação e ao acesso à informação na era digital.

[...] a cadeia conceitual que caracteriza a Ciência da Informação vai desde o dado à informação e conhecimento [...] à sua elaboração como informação”, assim, a sua absorção quando por vez relevante na estrutura cognitiva, transforma-se em conhecimento, onde incorporações dessas informações constroem também processo de cultura, “entre outras manifestações e produções e vivências do homem, individuais e coletivas. (Pinheiro, 2005, p. 40).

Os avanços tecnológicos no campo da comunicação e informação desencadearam a necessidade de novas estratégias na documentação e disseminação de informações diante das ferramentas da tecnologia digital. Essa tecnologia, considerada uma das tendências inovadoras da revolução contemporânea, tem influência direta na qualidade das informações e permeia o convívio e a interação com documentos digitais (Tocci *et al.*, 2007). Portanto, os novos ambientes de informação são entendidos como unidades culturais de infocomunicação, dispositivos e tecnologias capazes de transmitir e mediar conhecimento intercultural.

Não se busca aqui fazer uma apologia ao digital e às Tecnologias de Informação e Comunicação (TIC), mas é importante ressaltar que sua presença, embora nem sempre popularizada e acompanhada de desigualdades de acesso, enriquece e disponibiliza novos meios e recursos para a criação cultural e para a dinamização das relações sociais. Segundo Lévy (1993) esses avanços permeiam o cotidiano, as ciências, as tecnologias e, em grande parte, as informações que acessamos em plataformas virtuais. Isso significa que a convivência com a

documentação na *web* de dados é inevitável, considerando seu impacto na sociedade e a necessidade de adesão da CI a esse meio.

A documentação na *web* de dados exige uma abordagem específica, considerando a complexidade e a diversidade dos níveis informacionais presentes nesse ambiente (Meadows, 1999). Os processos de qualificação de dados na *web*, conforme apontado por Decarli e Tzagaraki (2006), desempenham um papel crucial na pesquisa, preservação e comunicação da informação. Ao organizar e tratar a informação de maneira acessível, é possível atender às necessidades de diferentes públicos e objetivos.

Nesse sentido, a CI desempenha um papel fundamental na compreensão e no estudo do comportamento dos algoritmos nos processos de automação dos motores de busca. Ao investigar a interação entre os algoritmos e a informação disponível na *web* de dados, é possível obter *insights* sobre como os sistemas automatizados de busca processam e fornecem acesso à informação. Isso contribui para o avanço do campo e para o desenvolvimento de estratégias mais eficientes na recuperação e organização da informação.

A compreensão do termo "algoritmo" é fundamental para a investigação do comportamento deles nos processos de automação dos motores de busca, sob a perspectiva da CI. Toscani e Veloso (2001) conceituam algoritmos como uma sequência de raciocínios, instruções ou operações que visam alcançar um objetivo específico. Essa sequência de passos deve ser finita e operada de maneira sistemática. Por sua vez, Menezes (2001) enfatiza que o algoritmo se justifica pelo resultado almejado, e sua complexidade aumenta à medida que são consideradas diferentes situações e cenários possíveis.

Assim como em códigos, a compreensão detalhada dos algoritmos requer a leitura minuciosa, linha por linha, a fim de entender como cada instrução contribui para atingir o objetivo final. Nesse contexto, é relevante explorar o comportamento dos algoritmos nos processos de automação dos motores de busca, uma vez que esses algoritmos desempenham um papel crucial na recuperação e organização da informação na *web* de dados. Compreender como os algoritmos operam, considerando a complexidade crescente e a adaptação a diferentes cenários, contribui para aprimorar a eficiência e a precisão dos motores de busca, bem como para otimizar a experiência dos usuários na busca de informações na era digital.

Portanto, investigar o comportamento dos algoritmos dos processos de automação dos motores de busca na perspectiva da CI implica em analisar cuidadosamente as sequências de instruções e operações que compõem esses algoritmos, considerando sua lógica, funcionamento e impacto na recuperação e organização da informação.

A compreensão desse estudo, sob a perspectiva da CI, está intrinsecamente relacionada às estruturas e elementos que compõem os algoritmos. Essas estruturas, como as variáveis que representam as informações de entrada e saída, desempenham um papel fundamental na determinação dos processos (Araújo, 2021; Mueller; Massaron, 2018). Ao compor sequências de algoritmos, os programas viabilizam a criação de *softwares* e facilitam os processos de busca, navegação e interação do usuário com os sistemas, permitindo a automação das tarefas.

Nesse contexto, é importante ressaltar que esta pesquisa busca estabelecer uma conexão com os estudos da CI relacionados à organização e recuperação da informação, com o objetivo de preencher lacunas de conhecimento sobre o tema e contribuir para a produção acadêmico-científica do Programa de Pós-Graduação em CI. A escolha dessa linha de pesquisa se apoia na compreensão da interface entre informação, tecnologia social e digital.

2.3 Relações Interdisciplinares envolvendo a Ciência da Informação e o estudo dos Algoritmos

Dessa forma, essa abordagem permite a percepção da dos algoritmos que, de forma implícita, abrangem os estudos e pesquisas relacionados às plataformas informacionais. Além disso, engloba a análise das competências informacionais, inclusão digital, comportamentos e hábitos informacionais.

Ao analisar o comportamento dos algoritmos dos processos de automação dos motores de busca, é essencial considerar a gestão do conhecimento e as tecnologias digitais como elementos intrínsecos a essa investigação. Dessa forma, gestão do conhecimento compreende a identificação, aquisição, estruturação e disseminação de informações, com o propósito de aperfeiçoar a tomada de decisões e estimular a inovação nas organizações. Nesse contexto, os algoritmos desempenham um papel crucial na busca, seleção e classificação das informações disponíveis, tornando possível a efetiva gestão do conhecimento.

Ademais, a tecnologia digital desempenha um papel fundamental na automação dos processos de busca e Recuperação de Informações (RI). Os motores de busca, impulsionados pelos algoritmos, são responsáveis por indexar, analisar e apresentar os resultados relevantes para os usuários, facilitando o acesso à informação em ambientes digitais (Baeza-Yates; Ribeiro-Neto, 1999). Compreender o comportamento dos algoritmos dos motores de busca na perspectiva da CI significa explorar as interações entre tecnologia, informação e sociedade, a

fim de promover uma melhor utilização e compreensão dos recursos informacionais disponíveis.

Nesse contexto, a análise visa compreender a interação dos usuários com os motores de busca e como suas necessidades de informação são supridas. Através dessas pesquisas, torna-se possível detectar eventuais lacunas e desafios na relação entre os usuários e a tecnologia, bem como elaborar estratégias que incentivem a utilização eficaz da informação e aprimorem a experiência do usuário.

Portanto, ao considerar o comportamento dos algoritmos dos processos de automação dos motores de busca na perspectiva da CI, é imprescindível abordar não apenas os aspectos técnicos e tecnológicos, mas também as dimensões sociais, culturais e organizacionais envolvidas. Essa abordagem multidisciplinar e holística permitirá uma compreensão mais abrangente dos impactos e implicações desses algoritmos na sociedade, contribuindo para o desenvolvimento de estratégias mais eficientes de organização, acesso e uso da informação em um contexto cada vez mais digital e globalizado.

Além disso, a compreensão do comportamento dos algoritmos dos processos de automação dos motores de busca na perspectiva da CI também se conecta com discussões sobre ética, privacidade e viés informacional. À medida em que os algoritmos assumem um papel cada vez mais central na organização e seleção das informações apresentadas aos usuários, surgem preocupações sobre a imparcialidade e a neutralidade desses sistemas automatizados (Noble, 2018).

Autores como Gillespie (2014) ressaltam a importância de examinar criticamente os algoritmos e suas implicações sociais, uma vez que esses mecanismos de filtragem e classificação não são neutros, mas sim influenciados por decisões humanas e por dinâmicas de poder. Através da análise do comportamento dos algoritmos dos motores de busca, é possível identificar possíveis vieses e discriminações presentes nos resultados apresentados, reforçando desigualdades existentes na sociedade (O'neil, 2016).

Outro aspecto relevante a ser considerado é a questão da privacidade dos usuários. A coleta e o processamento de dados pessoais pelos motores de busca e suas consequências para a proteção da privacidade têm sido objeto de debate (Zimmer, 2010). Autores como Zuboff (2019) destacam a emergência de uma nova forma de capitalismo de vigilância, no qual os dados pessoais dos usuários são utilizados como um recurso estratégico para a tomada de decisões comerciais e para a influência de comportamentos.

Ao aprofundar a compreensão do comportamento dos algoritmos dos motores de busca, é importante dialogar com as teorias e conceitos da CI. Autores como Saracevic (1999)

e Buckland (1991) exploram a natureza da informação, sua organização, recuperação e uso. A partir dessas perspectivas, é possível analisar como os algoritmos influenciam o processo de busca e seleção da informação, impactando diretamente a experiência do usuário.

Portanto, compreender o comportamento dos algoritmos dos processos de automação dos motores de busca na perspectiva da CI exige uma abordagem interdisciplinar, envolvendo áreas como ética, privacidade, viés informacional e teorias da informação. Essa análise crítica e aprofundada permitirá uma reflexão mais completa sobre o papel desses algoritmos na sociedade contemporânea e subsidiará o desenvolvimento de políticas e práticas informacionais mais justas e transparentes. De acordo com os princípios da CI, os algoritmos de pesquisa levam em conta uma série de elementos, tais como as palavras-chave da consulta, a relevância e usabilidade das páginas, a experiência das fontes de informação, bem como a localização e as configurações do usuário.

É importante destacar que o peso atribuído a cada um desses fatores pode variar, dependendo da natureza específica da consulta realizada. Nesse sentido, compreender de que forma os algoritmos operam nos processos de automação dos motores de busca do GS é um questionamento central para a pesquisa em CI, visando aprofundar o entendimento da interação entre usuários e sistemas de informação automatizados, bem como esclarecer a tomada de decisões informadas no contexto da sociedade contemporânea.

A pesquisa em CI relacionada ao comportamento dos algoritmos dos motores de busca encontra respaldo em diversos estudos e teorias que destacam a importância desse tema. Autores renomados contribuem para a compreensão dessas questões, como Xie e Li (2018) e Gonçalves e Fox (2019).

Xie e Li (2018) enfatizam que a análise aprofundada dos mecanismos de busca na *web* é fundamental para compreender como os algoritmos determinam a relevância dos resultados de pesquisa. Essa compreensão proporciona uma visão mais abrangente das estratégias de indexação e *ranking* adotadas, permitindo aprimorar a eficácia e a transparência desses algoritmos. Além disso, os autores ressaltam a necessidade de investigar os fatores e sinais considerados pelos algoritmos pela sua posição dominante no campo da tecnologia digital.

Gonçalves e Fox (2019), por sua vez, destacam a importância da implementação de *scripts* de busca automatizada para explorar a interação entre a automação dos processos de busca e os algoritmos utilizados. Essa abordagem prática permite uma compreensão mais aprofundada do fluxo de trabalho dos motores de busca, abrangendo desde a coleta de dados até a apresentação dos resultados. Os autores salientam a relevância da seleção de dados

relevantes, da eficiência do algoritmo e do gerenciamento de grandes volumes de informações nesse contexto.

Assim, ao realizar essa pesquisa, busca-se contribuir para o avanço do conhecimento em CI, embasando-se nas teorias e estudos de Xie e Li (2018) e Gonçalves e Fox (2019). Proporcionando uma base sólida de compreensão teórica e prática sobre o comportamento dos algoritmos dos processos de automação dos motores de busca, os *insights* obtidos têm o potencial de influenciar o desenvolvimento de estratégias de busca eficazes, o aprimoramento de sistemas de recuperação da informação e a garantia de acesso equitativo e confiável às fontes de informação na era digital.

Além dos autores mencionados e outros pesquisadores também têm contribuído significativamente para o entendimento do comportamento dos algoritmos dos processos de automação dos motores de busca. Por exemplo, Singh e Kant (2020) realizaram estudos sobre a importância da personalização dos resultados de busca e seus impactos na relevância e na qualidade das informações apresentadas aos usuários. Já Liu e Zhu (2021) abordaram a influência dos algoritmos de recomendação nos resultados de pesquisa e nas experiências dos usuários.

No contexto prático, a implementação de um *script* de busca automatizada, utilizando a linguagem *Python*¹ como exemplo de *Web Crawler*, permitirá explorar a interação entre a automação dos processos de busca e os algoritmos utilizados. Essa abordagem fornece uma compreensão mais aprofundada do fluxo de trabalho dos motores de busca, desde a coleta de dados até a apresentação dos resultados, enfatizando a importância da seleção de informações relevantes, da eficiência dos algoritmos e do gerenciamento de grandes volumes de dados.

Portanto, à medida que avançamos em direção ao desconhecido, inspirados por possibilidades ilimitadas, seguimos o exemplo de Solana-González e Pérez-González (2021) ao abraçar a exploração contínua da interseção entre tecnologia e conhecimento. Nesse espírito, a ampliação de nossa compreensão sobre algoritmos de busca e a busca por soluções que aprimorem a interação humana com a informação se alinham a um futuro de progresso e inovação constante.

¹ *Python* é uma linguagem de programação de alto nível amplamente utilizada em diversas áreas, desde desenvolvimento de aplicações *web* até automação e análise de dados. No campo científico, é amplamente adotado devido à sua capacidade de processar e analisar dados de forma eficiente. Bibliotecas como *NumPy*, *Pandas* e *Matplotlib* tornaram-se pilares para cientistas de dados, pesquisadores e engenheiros que buscam realizar análises estatísticas, criar visualizações de dados e trabalhar com grandes conjuntos de dados. Disponível em: <https://www.python.org/>. Acesso em: 8 set. 2023.

Certamente, a interseção entre a pesquisa sobre algoritmos de motores de busca e a CI é rica em contribuições de diversos autores, cujas perspectivas e reflexões enriquecem o campo. Autores como Batista (2014), Marchand e Davenport (2000), e Riemenschneider *et al.* (2002), entre outros, forneceram um arcabouço sólido para a compreensão do rigor metodológico, da gestão da informação e da relevância dos algoritmos de busca.

Batista (2014), com sua obra "Gestão da Informação: Como transformar informações em ativos para a empresa", trouxe à tona a importância da gestão informacional como um processo estratégico. Sua perspectiva de captura, organização e armazenamento de dados está alinhada com a essência dos algoritmos de busca, que buscam extrair, organizar e disponibilizar informações relevantes para os usuários. A pesquisa sobre algoritmos de busca pode potencializar essa abordagem, criando ferramentas mais eficazes para a gestão de informações nas organizações, sejam elas empresas, governos ou instituições sem fins lucrativos.

Marchand e Davenport (2000), por sua vez, ressaltaram a importância da interação colaborativa entre *stakeholders* e usuários finais na criação de soluções úteis. Sua abordagem alinha-se com a necessidade de compreender as necessidades e expectativas dos usuários ao projetar algoritmos de busca. A pesquisa nesta área pode aprimorar a interação entre humanos e sistemas, resultando em algoritmos mais eficazes e centrados no usuário.

Riemenschneider *et al.* (2002) destacaram o conceito de rigor como fundamental na investigação científica, enfatizando a consistência metodológica e a aderência a princípios de pesquisa rigorosos. Isso se reflete diretamente na pesquisa sobre algoritmos de busca, onde a precisão na análise, a validação metódica dos resultados e a consideração cuidadosa de fatores éticos e técnicos são vitais.

Ao promover o avanço do conhecimento em CI, essa pesquisa visa contribuir para o desenvolvimento de estratégias de busca mais eficazes, o aprimoramento dos sistemas de recuperação da informação e a garantia de acesso equitativo e confiável às fontes de informação na era digital.

3 ALGORITMOS DE BUSCA NA WEB

A pesquisa acerca do comportamento dos algoritmos de busca na *web*, no contexto da automatização dos motores de busca, desempenha um papel de suma importância no campo da CI. Por meio desta investigação, almeja-se alcançar uma compreensão aprofundada desses algoritmos, que desempenham um papel crítico na otimização das pesquisas realizadas na *internet*, independentemente da natureza do conteúdo almejado. Os motores de busca representam ferramentas de valor inestimável na facilitação do acesso à informação na era digital, viabilizando a localização de resultados pertinentes dentre o vasto volume de dados disponíveis.

Os algoritmos de busca assumem a responsabilidade de rastrear e indexar a *web*, empregando aranhas ou *bots* automatizados para percorrer a *internet* e extrair informações dos *websites*. Estas informações são posteriormente organizadas e categorizadas em um índice, o qual constitui a base para a apresentação dos resultados de pesquisa aos usuários. Nesse sentido, a eficácia do processo de rastreamento e indexação exerce um papel de importância crucial na classificação e visibilidade dos *websites* nos resultados de busca.

A compreensão dos fatores e sinais considerados pelos algoritmos de busca é fundamental, pois engloba critérios de relevância, padrões de classificação e técnicas de indexação adotadas, oferecendo uma visão holística das estratégias empregadas pelos motores de busca. Estes *insights* são de suma importância para aprimorar a eficácia e transparência dos algoritmos, garantindo que os resultados de pesquisa alcancem um nível crescente de relevância e confiabilidade.

Ademais, a implementação de um exemplo prático de *Web Crawler*, por meio da linguagem de programação *Python*, proporciona uma oportunidade para explorar a interação intrínseca entre os processos de automação de busca e os algoritmos aplicados. Essa abordagem possibilita uma compreensão mais profunda do fluxo de trabalho dos motores de busca, desde a fase de coleta de dados até a apresentação dos resultados. Ao analisar o funcionamento do *Web Crawler* e os dados que ele reúne, torna-se viável identificar padrões, tendências e desafios associados aos algoritmos de busca, contribuindo, assim, para o progresso do conhecimento no âmbito da CI. Com base nas questões problemáticas previamente delineadas, esta pesquisa tem como propósito abordar e desenvolver uma compreensão aprofundada dos algoritmos nos processos de automação dos motores de busca, sob a perspectiva da CI. Os algoritmos empregados por esses motores de busca desempenham uma função crucial ao navegar pela

vasta rede mundial de maneira metódica e automatizada. Sua responsabilidade reside em encontrar, analisar e indexar páginas de um *site* ou informações em um sistema específico.

Pode-se considerar esses algoritmos como "agentes virtuais" que capturam informações a partir dos *links* disponíveis na *internet*, catalogando e discernindo o que é de maior relevância. Esta abordagem possibilita a análise do código de um *website* a fim de buscar informações e gerar entendimento.

A compreensão detalhada dos algoritmos nos processos de automação dos motores de busca, sob a ótica da CI, ostenta grande relevância devido ao papel crucial desempenhado por esses algoritmos na indexação, organização e disseminação de informações.

A pesquisa destes ciclos, processos, hábitos e comportamentos informacionais, que englobam as atividades de leitura, escrita e circulação da informação, exige a exploração dos mecanismos subjacentes a esta ferramenta, proporcionando uma experiência multidisciplinar que abrange diversas áreas de conhecimento, incluindo a própria CI. A interligação entre a CI e o universo digital, em particular os algoritmos, revela-se imprescindível no contexto da automação no cenário infodigital. A compreensão aprofundada do comportamento dos algoritmos de rastreamento e indexação na *web*, integrados aos processos de automação dos motores de busca, destaca-se como um pilar fundamental no campo da CI. Ao explorar os mecanismos de busca na *web*, impulsionados por tais algoritmos, é possível obter respostas mais eficientes e precisas quando comparadas a alternativas disponíveis. Nesse contexto, o principal objetivo desta pesquisa reside em fornecer uma base sólida de conhecimento que capacite os usuários a compreender a funcionalidade e a tecnologia subjacente a esses algoritmos, especialmente no que tange aos processos de automação, rastreamento e indexação *web*.

A automação está se tornando cada vez mais presente nos processos informacionais, abrangendo desde tarefas simples até as mais complexas. Nesse contexto, é fundamental compreender a relevância dos algoritmos de rastreamento e indexação *web* na CI, reconhecendo-os como instrumentos facilitadores na busca por informação e conhecimento. Além disso, eles desempenham um papel crucial na análise, coleta, classificação, manipulação, armazenamento, recuperação e disseminação de informações na *web*. Assim, esses algoritmos asseguram que a informação presente na *web* seja assimilada desde seu surgimento até o processo de transformação de dados em conhecimento.

A compreensão do comportamento dos algoritmos de automação dos motores de busca, na perspectiva da CI, é de extrema importância, considerando o papel fundamental desses algoritmos na busca e organização de informações na *web*. Esses algoritmos, como

técnicas especializadas de serviços *on-line*, são desenvolvidos para realizar pesquisas em outros *sites*, com base nas consultas dos usuários. Ao procurar resultados em seu banco de dados, os mecanismos de busca classificam as páginas com base em critérios específicos e geram uma lista ordenada de resultados por meio de algoritmos de pesquisa exclusivos. Essa lista, conhecida como Resultados do Mecanismo de Pesquisa (*Search Engine Results Page*) (SERP), exhibe os resultados mais relevantes para uma determinada palavra-chave ou consulta.

A SERP desempenha um papel crucial na experiência do usuário, ao fornecer os resultados de uma consulta e servir como uma ponte para atender às necessidades informacionais. Utilizando algoritmos de rastreamento e indexação, como os *Web Spiders*, os mecanismos de busca conseguem localizar, ler e indexar as páginas dos *sites*, garantindo a entrega eficaz e precisa das informações desejadas pelos usuários. Esses algoritmos, ao identificar padrões nos dados coletados, contribuem para a interpretação de informações relevantes e auxiliam nas tomadas de decisão futuras.

Nesse sentido, a criação de um rastreador da *web* emerge como uma abordagem pertinente para localizar e coletar conteúdo completo ou parcial de *websites*, tornando o acesso à informação mais gerenciável. Os dados coletados podem ser armazenados em um banco de dados, integrados a um sistema interno ou adaptados a outras ferramentas. Além disso, há diversas maneiras de acessar e utilizar esses dados de maneira estruturada. A funcionalidade do rastreador pode ser expandida para realizar classificações do conteúdo, como a extração de dados em um período específico ou por palavras-chave, atendendo às necessidades específicas de cada pesquisa.

A relevância e justificativa desta pesquisa residem no papel crucial desempenhado pelos motores de busca na sociedade atual, caracterizada pela digitalização e informatização abrangentes. Esses motores representam um elemento conceitual fundamental para a RI em uma ampla variedade de domínios, incluindo acervos científicos, acadêmicos, literários e outros. Aprofundar a compreensão do comportamento dos algoritmos desses motores e explorar suas funcionalidades oferece oportunidades significativas para identificar novas perspectivas, aprimorar o acesso à informação e otimizar os processos de busca.

Além de beneficiar os usuários ao facilitar a obtenção de informações relevantes, esta pesquisa contribui substancialmente para o avanço do campo da CI. Ao proporcionar uma base sólida de conhecimento teórico e prático sobre os algoritmos dos motores de busca e suas implicações na sociedade contemporânea, ela promove uma compreensão mais profunda dos mecanismos subjacentes a essas ferramentas de busca. A justificativa para este estudo está na necessidade premente de compreender os algoritmos dos motores de busca, explorando sua

aplicação no âmbito da CI e reconhecendo sua relevância na automatização dos processos de busca por informações.

A interação entre a CI e os algoritmos dos motores de busca possibilita um avanço significativo no conhecimento, permitindo a otimização da RI, a compreensão dos fluxos informacionais e a construção de significado no contexto digital. A investigação e exploração desses aspectos contribuem de maneira fundamental para a melhoria dos sistemas de busca e para a eficácia na gestão da informação na sociedade contemporânea.

Além disso, é importante destacar que os algoritmos de busca não se limitam à democratização e descentralização da informação; eles se transformaram em verdadeiras bibliotecas virtuais, abrigando informações de uma ampla gama de gêneros e categorias. Nesse contexto, esta pesquisa oferece uma valiosa contribuição para a CI, especialmente no campo da gestão de informações, ao agregar e disponibilizar informações de maneira eficiente e acessível. Dessa forma, a investigação sobre os algoritmos dos motores de busca, no contexto da CI, revela-se essencial para a compreensão do comportamento dessas ferramentas, sua aplicação na RI e sua relevância na sociedade digital. Ao abordar esses aspectos, esta pesquisa contribui significativamente para o avanço do conhecimento na área e para a otimização dos processos de busca e gestão da informação.

Além disso, a revisão da literatura abordará as teorias que fundamentam a recuperação da informação e a relevância dos algoritmos dos motores de busca nesse contexto. Serão discutidos os modelos de recuperação da informação, como o modelo *Booleano*, o modelo Vetorial e o modelo Probabilístico, e como os Algoritmos de busca se relacionam com estes modelos.

O algoritmo possui estruturas como variáveis que são as informações de entrada e saídas que determinam os processos (Araujo, 2021; Mueller; Massaron, 2018). Nesse sentido, os programas são compostos por sequências de algoritmos que viabilizam todo tipo de *softwares* e que facilitam os processos de busca, navegação e interação do usuário com sistemas, permitindo a automação dos processos.

Os algoritmos podem ser definidos como sequências de instruções lógicas que resolvem um determinado problema ou executam uma tarefa específica (Cormen *et al.*, 2009, p. 5). Eles são amplamente utilizados na área da computação para processar dados e realizar operações complexas (Cormen *et al.*, 2009, p. 5). Além disso, os algoritmos têm aplicações em diversos domínios, abrangendo desde a matemática e a ciência da computação até "sistemas de busca na *web* e *mídias* sociais" (Cormen *et al.*, 2009, p. 5). Outras definições e autores que abordam o conceito de algoritmo incluem:

Knuth (1968), em seu livro "*The Art of Computer Programming*", define algoritmo como uma sequência de passos bem definidos e não ambíguos que levam a uma solução de um problema. Leiserson *et al.* (2009), em seu livro "*Introduction to Algorithms*", fornecem uma definição abrangente: Um algoritmo é um conjunto bem definido de regras que especifica uma sequência de operações a serem executadas para resolver um determinado problema em um número finito de etapas (Cormen *et al.*, 2009, p. 4). Robert Sedgewick e Kevin Wayne, em "*Algorithms*" (2011, p. 5), descrevem um algoritmo como uma sequência precisa de instruções bem definidas para executar uma tarefa ou resolver um problema.

O termo "algoritmo" deriva do sobrenome Al-Khwarizmi, do matemático persa do século IX, Mohamed ben Musa, cujas obras foram traduzidas no ocidente cristão no século XII (*The Editors Of Encyclopædia Britannica*). A palavra também tem derivações como "álgebra" e "algarismo" (Tavares, 1998). O conceito de algoritmo é frequentemente associado a uma lista de procedimentos bem definidos, podendo ser visualizado por meio de um fluxograma, representando a ideia da programação imperativa (Spitznagel, 1971).

Essas definições convergem em destacar que os algoritmos são sequências de instruções lógicas que visam resolver problemas ou realizar tarefas específicas. Eles são aplicáveis em diversos domínios, desde a computação em geral até sistemas específicos como os de busca na *web* e *mídias* sociais. A importância dos algoritmos reside na sua capacidade de processar dados e executar operações complexas, fornecendo soluções eficientes para os desafios encontrados em diferentes áreas de conhecimento.

Dessa forma, a contribuição de Salazar destaca a necessidade de reflexão sobre o papel dos algoritmos na formação de opiniões e na tomada de decisões, chamando a atenção para o poder que essas tecnologias possuem na influência da nossa percepção do mundo digital. Sua abordagem amplia o entendimento dos algoritmos além de suas funcionalidades técnicas, ressaltando sua dimensão sociocultural e ética na era da informação.

Os algoritmos podem perpetuar desigualdades e discriminação, pois são treinados em conjuntos de dados que podem estar enviesados ou conter estereótipos. Menciona, também, casos em que algoritmos de recrutamento, por exemplo, reproduziram desigualdades de gênero e raça, selecionando candidatos com base em padrões históricos de contratação. Esses exemplos destacam a importância de uma análise crítica dos algoritmos e a necessidade de desenvolver práticas que minimizem o viés e promovam a equidade.

A importância da transparência na concepção e implementação dos algoritmos, permite que os usuários compreendam como as decisões são tomadas e quais critérios estão

sendo aplicados. Além disso, há importância de mecanismos de prestação de contas e de avaliação contínua dos algoritmos para identificar e corrigir possíveis vies e discriminação.

Esses autores, juntamente a outros pesquisadores, têm contribuído para uma compreensão mais atualizada dos algoritmos e suas implicações. Argumentam que os algoritmos não são meramente sequências de instruções matemáticas, mas sistemas complexos que exercem influência sobre a cultura, a política, a economia e a sociedade como um todo. Eles fornecem uma perspectiva atualizada sobre o conceito de algoritmos, destacando seu papel como agentes de poder e influência na era digital. Suas abordagens têm sido estabelecidas por meio de análises críticas e reflexões sobre as implicações dos algoritmos em diferentes áreas, estimulando o debate e a conscientização sobre o tema.

A compreensão dos processos de automação da *web* é de fundamental importância no campo da CI. Nesse contexto, é necessário explorar a fundamentação teórica relacionada aos algoritmos de rastreamento, indexação e *ranking* utilizados na busca e extração de dados na *web*. Diversos autores contribuem para essa discussão, fornecendo compreensão sobre as ferramentas e metodologias empregadas nesses processos, bem como abordando as questões éticas e de privacidade envolvidas.

Freire (2002) destaca o papel da CI como uma disciplina interdisciplinar que promove a responsabilidade social do pensamento científico. Ele enfatiza que a CI facilita o acesso e a comunicação do conhecimento produzido, reconhecendo a informação como uma poderosa força de transformação social. Essa perspectiva ressalta a importância de compreender os algoritmos utilizados na *web*, pois eles desempenham um papel fundamental na organização e disseminação da informação.

Ribeiro e Santos (2020) fornecem uma definição abrangente de informação, considerando-a como dados que receberam significado e contextualização por um agente. Essa definição destaca a importância da extração e indexação dos dados na *web*, uma vez que os algoritmos desempenham um papel crucial na transformação desses dados brutos em informações relevantes e acessíveis.

Ao revisar a literatura existente, diversos autores também contribuem para o entendimento dessas questões. Araújo (2018), Ortega e Lara (2010), Rayward (1997), Albagli (2015), Silva e Freire (2012), Ceravolo e Tálamo (2000), Meadows (1999), Pinheiro (2005), Tocci (2007), Lévy (1993), Toscani e Veloso (2001), Mueller (2018) e Becker (1999) oferecem perspectivas únicas sobre os algoritmos e sua relação com a CI.

Esses autores abordam diversos aspectos, como a importância dos algoritmos na classificação e organização da informação, a influência dos algoritmos na descoberta de novos

conhecimentos, o papel da Inteligência Artificial (IA) na extração de informações relevantes e a interação entre os usuários e os sistemas de busca baseados em algoritmos. Suas contribuições fornecem uma base sólida para a compreensão dos princípios e conceitos subjacentes aos algoritmos na *web*.

Ao explorar os diversos aspectos relacionados à importância dos algoritmos na classificação e organização da informação, encontramos contribuições significativas de autores como Silva e Freire (2012) e Ceravolo e Tálamo (2000). Silva e Freire destacam que os algoritmos desempenham um papel central na organização dos resultados de busca, permitindo a classificação e a estruturação da informação de forma a facilitar sua recuperação e compreensão pelos usuários. Segundo eles, os algoritmos são responsáveis por filtrar e ordenar os resultados com base em critérios como relevância, popularidade e autoridade, o que contribui para melhorar a experiência de busca dos usuários.

Já Ceravolo e Tálamo (2000) enfatizam a influência dos algoritmos na descoberta de novos conhecimentos. Eles argumentam que os algoritmos de busca, ao selecionarem e apresentarem informações relevantes aos usuários, têm o potencial de ampliar o acesso a diferentes perspectivas e conhecimentos. Por meio desses algoritmos, é possível explorar uma variedade de fontes e descobrir informações que podem ser valiosas para a geração de novas ideias e entendimento.

No que diz respeito ao papel da IA na extração de informações relevantes, autores como Araújo (2018) e Mueller (2018) contribuem para a compreensão desse aspecto. Araújo (2018) destaca que os algoritmos de IA são capazes de processar grandes volumes de dados e identificar padrões e relações que seriam difíceis de serem percebidos pelos seres humanos. Essa capacidade de análise automatizada permite extrair informações relevantes que podem subsidiar tomadas de decisão e estudos científicos.

Mueller (2018), por sua vez, explora a interação entre os usuários e os sistemas de busca baseados em algoritmos. Ele argumenta que os algoritmos atuam como intermediários entre os usuários e a informação disponível na *web*, filtrando e apresentando os resultados de busca de acordo com as características e preferências dos usuários. Essa interação entre os usuários e os algoritmos é dinâmica e influencia o processo de busca, uma vez que os usuários podem fornecer *feedback* e interagir com os resultados apresentados, contribuindo para a melhoria contínua dos algoritmos.

Essas contribuições dos autores fornecem uma base sólida para a compreensão dos princípios e conceitos subjacentes aos algoritmos na *web*. Ao considerar a importância dos algoritmos na classificação e organização da informação, a influência na descoberta de novos

conhecimentos, o papel da IA na extração de informações relevantes e a interação com os usuários, é possível perceber a complexidade e a relevância desses processos na CI e na sociedade como um todo.

A partir dessas discussões, podemos compreender que os algoritmos desempenham um papel fundamental na *web*, especialmente no contexto da CI. Eles possibilitam a organização e a estruturação dos dados disponíveis, facilitando a recuperação e o acesso à informação relevante para os usuários.

A importância dos algoritmos na classificação e organização da informação é destacada por Silva e Freire (2012). Segundo eles, os algoritmos permitem a criação de sistemas de busca eficientes, capazes de filtrar grandes volumes de dados e apresentar resultados relevantes com base em critérios estabelecidos. Essa capacidade de classificação e organização contribui para a otimização dos processos de busca e recuperação de informação, tornando-os mais eficazes e precisos.

Além disso, os algoritmos também desempenham um papel significativo na descoberta de novos conhecimentos. Ceravolo e Tálamo (2000) ressaltam que, por meio dos algoritmos de busca, os usuários têm a oportunidade de acessar diferentes perspectivas e fontes de informação, ampliando suas possibilidades de descoberta e aprendizado. Os algoritmos selecionam e apresentam conteúdos relevantes, permitindo que os usuários explorem novas ideias e conhecimentos em diversas áreas do conhecimento.

A IA, por sua vez, desempenha um papel cada vez mais importante na extração de informações relevantes. Araújo (2018) destaca que os algoritmos de IA são capazes de analisar grandes conjuntos de dados de forma automatizada, identificando padrões e relações que podem ser relevantes para os usuários. Essa capacidade de processamento e análise de dados em larga escala possibilita a extração de informações que podem ser aplicadas em diferentes contextos, incluindo a pesquisa científica.

Por fim, a interação entre os usuários e os sistemas de busca baseados em algoritmos é um aspecto essencial a ser considerado. Mueller (2018) destaca que os usuários fornecem respostas constantes para os sistemas de busca, influenciando os algoritmos e refinando os resultados apresentados. Essa interação dinâmica permite uma melhoria contínua dos algoritmos, adaptando-os às necessidades e preferências dos usuários e garantindo uma experiência de busca mais personalizada e relevante.

Em suma, a compreensão dos princípios e conceitos subjacentes aos algoritmos na *web* é de extrema importância no campo da CI. Os estudos e abordagens dos autores mencionados proporcionam uma base sólida para explorar a importância dos algoritmos na

classificação e organização da informação, sua influência na descoberta de novos conhecimentos, o papel da IA na extração de informações relevantes e a interação entre os usuários e os sistemas de busca. Essas discussões contribuem para uma compreensão mais aprofundada dos processos de busca e recuperação de informação na era digital.

Na sequência, será realizada uma análise aprofundada dos algoritmos dos motores de busca. Serão discutidas as características e funcionalidades desses algoritmos, bem como sua relevância na recuperação da informação e na otimização dos processos de busca. Como também, exploradas as abordagens teóricas e os estudos existentes sobre os algoritmos de busca, destacando suas implicações na sociedade atual e os desafios relacionados à sua eficácia e transparência.

3.1 Algoritmos e suas Estruturas

É inimaginável a quantidade de informações registradas nos últimos anos. São dados extraídos em diversos formatos e traduzido em informações com diversas finalidades, como texto, imagem, vídeo, áudio, e proveniente de diversas tecnologias, de redes sociais ao cartão de crédito, de *Internet* das Coisas (IoT) a Sistema Operacional (SO), de drones aos mais diversos tipos de sensores. Para tanto, o fluxo de informações que circulam nos últimos anos alcançou um patamar nunca imaginado e progride de maneira espantosa, deixam um marco histórico na contemporaneidade. De modo que o registro dessas informações se transforma radicalmente na passagem do analógico para o digital.

Neste momento histórico, a chamada era digital, têm nas mãos o poder da informação muito mais acessível do que há 20 anos. Por esta razão, são chamados de sociedade da informação, que aprende na prática a lidar com a segurança de dados. Deveras, as tecnologias facilitam a identificação e a captação de informações, tornam-se essenciais em um universo cada vez mais conectado e exigente quanto à conformidade informacional. Entretanto, as instituições públicas e privadas investem em segurança de dados, visando não apenas o crescimento, mas também a reputação confiável diante do cenário em que atuam, passando a garantir, com responsabilidade, a privacidade das informações de seus clientes e/ou indivíduos.

A propriedade corporativa e individual, de acordo com Zuboff, (2019) é um desafio que envolve a proteção de dados e a educação dos usuários em relação à segurança das informações. Pois seus dispositivos transmitem e armazenam, conhecido como capitalismo de vigilância, estes conceitos. Zuboff, (2019) complementa que referem-se ao exercício de tratamento de informações pessoais de forma comercial. Portanto, a regulação da propriedade

dos dados é uma das questões contemporâneas mais complexas que a sociedade enfrenta, em face do tratamento indiscriminado de informações pessoais, por isso, a importância de algoritmos no processo de desenvolvimento de equipamentos responsáveis pela geração, armazenamento, distribuição, ativação e recuperação apropriada.

Assim, algoritmos cada vez mais sofisticados são usados para encontrar padrões nos dados dos indivíduos que indiquem o que gostam de consumir, os esportes que praticam, com quem se relacionam e até decisões que tomarão no futuro. Esses algoritmos direcionam o sistema de sugestões e os filtros de conteúdo observados no uso da *internet* e de aplicativos que são exemplos do dia a dia, em que atuam na tentativa de influenciar as escolhas individuais. De modo que os dados são coletados, extraídos, analisados, processados e tratados por algoritmos simples ou inteligentes que analisam, comparam e tratam informações dos mais variados formatos a todo momento (Tutt, 2016).

Ao contrário do relato frequentemente reproduzido de Goldstine (1997) em *A History of Numerical Analysis*, as primeiras formações da análise numérica como disciplina surgiram muito antes do final da Segunda Guerra Mundial, das contribuições de John von Neumann ou do nascimento do computador digital. O foco estreito de Goldstine (1997) nos séculos XVI a XVIII e nos maiores nomes da matemática foi criticado por Folta (1988) e, mais recentemente, por Tournès (2009). Ambos apontaram que práticas computacionais, como agora codificadas em análise numérica, podem ser encontradas ao longo dos séculos, mas muitas vezes não em matemática “pura”, e sim em campos de matemática prática, mista ou aplicada.

Uma história da análise numérica deve levar em conta esse aspecto. Como disciplina, a análise numérica deveria emergir no final do século XIX e ter um novo impulso de disciplinarização após a Primeira Guerra Mundial. De fato, a Primeira Guerra Mundial e suas necessidades computacionais, especialmente em balística, parecem ter estimulado uma primeira revisão sistemática da literatura matemática a fim de encontrar métodos eficientes de computação (Aubin; Goldstein, 2014; Gluchoff, 2011). Isso levou à publicação dos primeiros manuais de cálculo prático durante o período entre guerras.

Nesse contexto, é importante destacar a relevância da dimensão coletiva do algoritmo no tratamento das informações, que impacta todo universo digital. Algoritmo é o nome dado à sequência de ações indicando exatamente o que o computador deve fazer para realizar uma tarefa ou resolver um problema. O termo algoritmo tem sua origem com o matemático do século IX *Muhammad ibn Mūsā al-Khwārizmī*, cujo sobrenome foi empregar na expressão latina para *Algoritmi*. O conceito por trás do nome existe há séculos. Os grandes matemáticos gregos, já usavam para descrever como encontrar números primos ou o maior

divisor comum entre dois números inteiros (Knuth, 1968). Algoritmo é o nome dado à sequência de instruções indicando exatamente o que o computador deve fazer para realizar uma tarefa ou resolver um problema.

Suas operações devem ser simples no processo da execução, e as ações complexas devem ser subdivididas em ações menores. Para isso, cada operação deve ser entendida com precisão para não haver margem de ambiguidades ou dúvidas. Sua sequência de ações deve ser finita, ou seja, o algoritmo deve, obrigatoriamente, ter um fim (Leite, 2006). Quando se fala de algoritmos para a programação de *softwares* e/ou *hardware* é imprescindível que ele seja eficiente e preciso com as instruções de forma clara, bem definida e não ambígua. Um bom algoritmo possui algumas características, dentre as quais destacamos:

- Ele necessariamente precisa ser finito na execução de um determinado problema;
- Não deve ter ambiguidade – sem espaço para dupla interpretação;
- Possuir a capacidade de receber dados de entrada do mundo real;
- Obter a potencialidade de gerar informações de saída para o mundo real;
- Efetivar as etapas descritas e torná-las alcançáveis em tempo finito.

Um algoritmo não se limita a um programa de computador, mas sim aos passos necessários para realizar uma tarefa. Sua implementação pode ocorrer em diferentes tipos de autômatos, como computadores ou até mesmo seres humanos. Diferentes algoritmos podem realizar a mesma tarefa usando diferentes instruções e requerendo mais ou menos tempo, espaço ou esforço (Cormen, 2010). A complexidade computacional de um algoritmo depende das estruturas de dados utilizadas. Algoritmos que realizam a mesma tarefa podem ter diferentes níveis de complexidade, influenciados pela escolha das estruturas de dados adequadas (Ziviani, 2011, p. 1). Por exemplo, algoritmos para se vestir podem ter diferentes sequências de ações, como vestir meias antes ou depois das calças, resultando em diferentes graus de dificuldade na execução (*Algorithmics*).

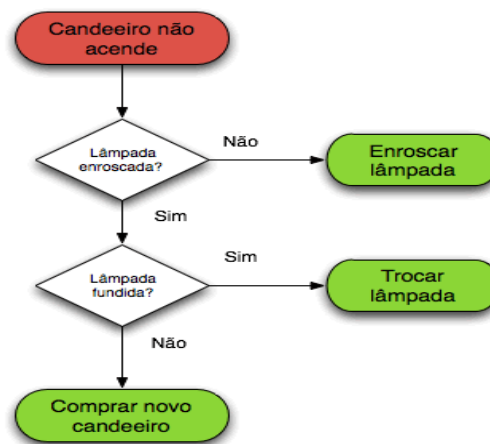
Um programa de computador pode ser entendido como um algoritmo que especifica ao computador os passos precisos e a ordem em que devem ser executados. Por exemplo, ele pode descrever os passos necessários para calcular as notas que serão impressas nos boletins dos alunos de uma escola. Portanto, um algoritmo pode ser considerado uma sequência de operações que podem ser simuladas por uma máquina de Turing completa (Ziviani, 2011, p. 1).

Quando um algoritmo envolve o processamento de dados, a informação é lida de uma fonte de entrada, processada e retornada com um novo valor após o processamento.

Geralmente, esse processamento é realizado com o auxílio de uma ou mais estruturas de dados (Loureiro, 2010) (Figura 1).

Para que um processo computacional seja eficiente e confiável, o algoritmo deve ser rigorosamente definido, especificando como ele se comportará em todas as circunstâncias. A corretividade do algoritmo pode ser provada matematicamente, assim como a análise da quantidade assintótica de tempo e espaço (complexidade) necessários para sua execução. Esses aspectos são investigados no campo da análise de algoritmos (Dasgupta; Papadimitriou; Vazirani, 2010, p. 2).

Figura 1 – Fluxograma Estrutural dos Algoritmos.



Fonte: Elaborado por Ocante António Ié (2023), baseado em Alscardoso (2018).

Uma forma comum de representar um algoritmo é por meio de uma lista de procedimentos bem definidos, na qual as instruções são executadas passo a passo a partir do início da lista. Essa representação pode ser visualizada de maneira clara através de um fluxograma, que ilustra graficamente o fluxo de controle do algoritmo (Mueller; Massaron, 2017). Essa abordagem é baseada nas premissas da programação imperativa, que fornece uma visão mecânica para desenvolver e compreender algoritmos. No entanto, existem outras concepções alternativas para algoritmos, como a programação funcional e a programação lógica (Cormen, 2010).

Quanto ao término de um algoritmo, alguns autores restringem sua definição a procedimentos que eventualmente terminam. Marvin Minsky observou que determinar antecipadamente o tamanho de um procedimento é um problema indecidível, uma vez que o procedimento pode ser executado infinitamente, impossibilitando a obtenção de uma resposta (Spitznagel, 1971). Alan Turing demonstrou em 1936 que não existe uma máquina de Turing capaz de resolver esse problema para todos os casos, o que ficou conhecido como o problema

da parada. Portanto, para algoritmos que não terminam, o sucesso não pode ser determinado apenas pela interpretação da resposta, mas depende das condições estabelecidas pelo desenvolvedor durante a execução (Spitznagel, 1971).

Essas características são evidentemente uma forma de explicar o funcionamento de algoritmos, em especial o da busca na *web*, como ele pode influenciar as preferências dos usuários. Ao longo da história da computação, vários algoritmos foram propostos para abordar questões de pesquisas nas plataformas *web* ou nos *softwares* específicos, levando ao método de IA, visto como solução de problemas por meio de busca.

Para Erickson (2019) os algoritmos de busca aparecem em diversas formas e são desenvolvidos para abordar a classe de problemas, que se tenta solucionar ou encontrar referências de soluções associadas. Nesse sentido, esses algoritmos trazem mudanças profundas no dia a dia da sociedade em geral, com sua dependência de motores de busca para fazer pesquisas.

3.2 Funcionamento de Motores de Buscas

Um mecanismo de busca é a aplicação prática de técnicas de RI para coleções de texto em larga escala. Um mecanismo de pesquisa na *web* é o exemplo óbvio, mas, como foi mencionado, os mecanismos de pesquisa podem ser encontrados em muitos aplicativos diferentes, como pesquisa na área de trabalho ou pesquisa corporativa. De acordo com Croft *et al.* (2015), os mecanismos de pesquisa existem há muitos anos. O termo “mecanismo de busca” foi originalmente usado para se referir a *hardware* especializado para busca de texto. A partir de meados da década de 1980, no entanto, gradualmente passou a ser usado em vez de “Sistema de Recuperação de Informações” como o nome do sistema de *software* que compara consultas a documentos e produz lista de resultados classificados de documentos.

As raízes da tecnologia do mecanismo de busca na *web* estão nos sistemas de RI, que remontam ao trabalho de Luhn na IBM durante o final dos anos 1950. A RI tem sido um campo ativo dentro da CI desde então, e recebeu um grande impulso desde a década de 1990 com os novos requisitos que a *web* trouxe. Muitos dos métodos usados pelos mecanismos de busca atuais podem ser rastreados até os desenvolvimentos em RI durante as décadas de 1970 e 1980. Especialmente influente é o sistema de recuperação Sistema para Análise Mecânica e Recuperação de Texto (SMART), inicialmente desenvolvido por Gerard Salton e seus colaboradores na Cornell *University* durante o início dos anos 1970 (Levene, 2010).

No início de 1999, o mecanismo de busca do *Google* começou a ganhar popularidade com seu novo paradigma de classificação baseado em *links* e uma *interface* de usuário simples. Em 2002 e 2003, o *Yahoo!* adquiriu a *Inktomi search* e a *Overture*, dona da *AltaVista* e da *AlltheWeb*, construindo sua própria tecnologia de busca. A *Ask.com* (anteriormente chamada de *Ask Jeeves*) adquiriu a pesquisa *Teoma* em 2001 para fornecer pesquisa e respostas a perguntas. A *Microsoft* usou inicialmente os resultados de pesquisa do *Inktomi* e gradualmente construiu seu próprio mecanismo: pesquisa do MSN, *Windows Live Search*, *Live Search* e *Bing*. Na Europa, o *Fast search* foi lançado em 1999 para competir com o *Google* e foi adquirido parcialmente pela *Overture* em 2003 e parcialmente pela *Microsoft* em 2007. Na China, o *Baidu* foi iniciado em 1999 e se tornou um mecanismo de busca dominante. Na Rússia, a pesquisa *Yandex* foi lançada em 1998 e atualmente está dominando, enquanto na Coreia do Sul, o *Naver* é um mecanismo de pesquisa popular lançado em 2000 (Levene, 2010).

A tecnologia por trás dos principais mecanismos de busca raramente é publicada devido à concorrência acirrada, enquanto algumas de suas técnicas têm aparecido gradualmente em patentes e conferências acadêmicas. Conferências acadêmicas que abordam mais sobre algoritmos de busca e análise de dados incluem *Special Interest Group on Information Retrieval Information Retrieval* (SIGIR), *World Wide Web* (WWW), *Web Search and Data Mining* (WSDM) e *Educational Resources Information Center* (ERIC). Outras conferências relacionadas para análise de dados da *web* incluem *Knowledge Discovery in Databases* (KDD), (*Conference on Information and Knowledge Management* (CIKM), *Special Interest Group on Management of Data* (SIGMOD) e *Very Large Database* (VLDB). A tecnologia de sistema que abrange sistemas de *Internet* de grande escala e plataformas de dados apareceu em várias conferências de sistema, como a *Ocular Surface Disease Index* (OSDI). Os mecanismos de pesquisa têm um grande impacto no tráfego de *sites*, e a atividade da indústria de pesquisa é observada de perto por *webmasters* e várias organizações, incluindo *searchenginewatch.com*. Livros de texto recentes que descrevem algoritmos de busca e técnicas de mineração podem ser encontrados em (Croft *et al.* 2015).

Criar um mecanismo de pesquisa que se adapte à *web* atual apresenta muitos desafios. A tecnologia de rastreamento rápido é necessária para reunir os documentos da *web* e mantê-los atualizados. O espaço de armazenamento deve ser usado de forma eficiente para armazenar índices e, opcionalmente, os próprios documentos. O sistema de indexação deve processar centenas de *gigabytes* de dados com eficiência. As consultas devem ser tratadas rapidamente, a uma taxa de centenas a milhares por segundo.

Essas tarefas estão se tornando cada vez mais difíceis à medida que a *web* cresce. No entanto, o desempenho e o custo do *hardware* melhoraram para compensar parcialmente a dificuldade. Existem, no entanto, várias exceções notáveis a esse progresso, como tempo de busca de disco e robustez do SO. O *Google* foi projetado para suportar conjuntos de dados extremamente grandes, fazendo uso eficiente do espaço de armazenamento para armazenar o índice, sendo suas estruturas de dados otimizadas para acesso rápido e eficiente.

De acordo o *Google* (2022), a pesquisa do *Google* é um mecanismo de pesquisa totalmente automatizado que usa um *software* conhecido como rastreadores que exploram a *web* regularmente para encontrar páginas e adicionar ao seu índice. Na verdade, a grande maioria das páginas listadas em seus resultados não são enviadas manualmente para inclusão, mas são encontradas e adicionadas automaticamente quando os rastreadores da *web* exploram a *web*.

A Pesquisa do *Google* funciona em três estágios, e nem todas as páginas passam por cada estágio:

- Rastreamento: o *Google* baixa textos, imagens e vídeos de páginas encontradas na *Internet* com programas automatizados chamados rastreadores.

A primeira etapa é descobrir quais páginas existem na *web*. Não há um registro central de todas as páginas da *web*, então o *Google* deve procurar constantemente por páginas novas, atualizadas e adicioná-las à sua lista de páginas conhecidas. Esse processo é chamado de "descoberta de URL". Algumas páginas são conhecidas porque o *Google* já as visitou. Outras páginas são descobertas quando o *Google* segue um *link* de uma página conhecida para uma nova página, por exemplo: uma página de *hub*, como uma página de categoria, *links* para uma nova postagem de *blog*. Ainda outras páginas são descobertas quando você envia uma lista de páginas (um mapa do *site*) para o *Google* rastrear.

Depois que o *Google* descobre o *Uniform Resource Locator* (URL) de uma página, ele pode visitar (ou "rastrear") a página para descobrir o que há nela, usando um conjunto de computadores para rastrear bilhões de páginas na *web*. O programa que faz a busca é chamado de *Googlebot* (também conhecido como rastreador, robô, *bot* ou *spider*). O *Googlebot* usa um processo algorítmico para determinar quais sites rastrear, com que frequência e quantas páginas buscar em cada *site*. Os rastreadores do *Google* também são programados para tentar não rastrear o *site* muito rápido para evitar sobrecarregá-lo. Esse mecanismo é baseado nas respostas do *site* (por exemplo, erros *Hypertext Transfer Protocol* (HTTP) 500 significam "desaceleração") e nas configurações do *Search Console*.

No entanto, o *Googlebot* não rastreia todas as páginas que descobriu. Algumas páginas podem não ter permissão para rastreamento pelo proprietário do *site*, outras páginas podem não estar acessíveis sem fazer *login* no *site*.

Durante o rastreamento, o *Google* renderiza a página e executa qualquer *JavaScript* que encontra usando uma versão recente do *Chrome*, semelhante à forma como seu navegador renderiza as páginas que você visita. A renderização é importante porque os *sites* geralmente dependem do *JavaScript* para trazer o conteúdo para a página e, sem a renderização, o *Google* pode não ver esse conteúdo. O rastreamento depende se os rastreadores do *Google* podem acessar o *site*. Alguns problemas comuns com o *Googlebot* acessando sites incluem:

- Problemas com o servidor que gerencia o *site*;
- Problemas de rede;
- Diretivas *robots.txt* que impedem o acesso do *Googlebot* à página.
- Indexação: o *Google* analisa os arquivos de texto, imagens e vídeos da página e armazena as informações no índice do *Google*, que é um grande banco de dados.

Depois que uma página é rastreada, o *Google* tenta entender do que se trata a página. Esse estágio é chamado de indexação e inclui o processamento e a análise do conteúdo textual e das principais marcas e atributos do conteúdo, como elementos *title* e atributos alternativos, imagens, vídeos e muito mais.

Durante o processo de indexação, o *Google* determina se uma página é uma duplicata de outra página na *internet* ou canônica. O canônico é a página que pode ser exibida nos resultados da pesquisa. Para selecionar o canônico, primeiro são agrupadas as páginas encontradas na *internet* com conteúdo semelhante e, em seguida, são selecionadas aquela que é mais representativa do grupo. As outras páginas do grupo são versões alternativas que podem ser exibidas em diferentes contextos, como se o usuário estiver pesquisando em um dispositivo móvel ou procurando uma página muito específica desse *cluster*.

O *Google* também coleta sinais sobre a página canônica e seu conteúdo, que podem ser usados na próxima etapa, onde veiculamos a página nos resultados de pesquisa. Alguns sinais incluem o idioma da página, o país em que o conteúdo é local, a usabilidade da página e assim por diante.

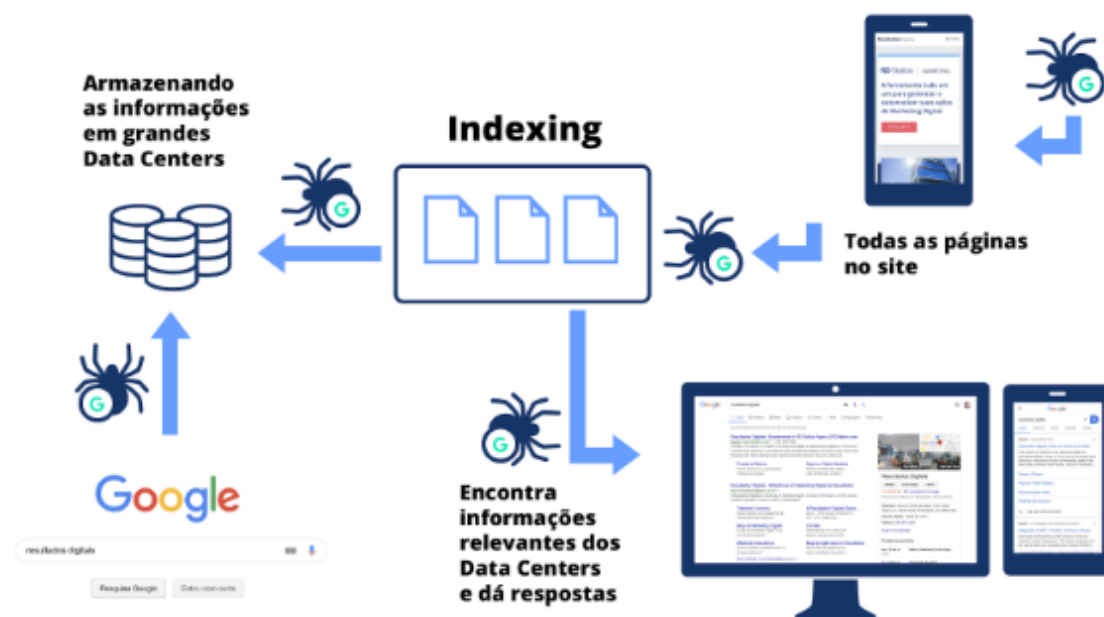
As informações coletadas sobre a página canônica e seu *cluster* podem ser armazenadas no índice do *Google*, um grande banco de dados hospedado em milhares de

computadores. A indexação não é garantida e nem todas as páginas processadas pelo *Google* serão indexadas.

A indexação também depende do conteúdo da página e de seus metadados (Figura 2). Alguns problemas comuns de indexação podem incluir:

- A qualidade do conteúdo da página é baixa;
- Metadiretivas de robôs não permitem indexação;
- O *design* do *site* pode dificultar a indexação.
- Exibição de resultados de pesquisa: quando um usuário pesquisa, o *Google* retorna informações relevantes para a consulta do usuário.

Figura 2 - Comportamento de Motores de Buscas.



Fonte: <https://encurtador.com.br/guyNQ> (2023).

Quando um usuário insere uma consulta, as máquinas do *Google* pesquisam o índice em busca de páginas correspondentes e retornam os resultados que acreditam serem da mais alta qualidade e mais relevantes para a consulta do usuário. A relevância é determinada por centenas de fatores, que podem incluir informações como localização, idioma e dispositivo do usuário (*desktop* ou *telefone*).

No âmbito da pesquisa *on-line*, é inegável a relevância dos motores de busca como ferramentas primordiais para a RI na internet. Dentre esses motores de busca, destaca-se o *Google*, cujo funcionamento tem sido objeto de amplo estudo e análise (Ribeiro *et al.*, 2020).

Neste contexto, é válido ressaltar que o presente estudo utiliza o *Google* como exemplo para ilustrar o funcionamento de um motor de busca, embora os mecanismos de buscas de *Google* e outras plataformas, não sejam o foco central da pesquisa.

Durante o processo de indexação, uma das etapas cruciais para o funcionamento eficiente dos motores de busca, é que, o *Google* estabelece critérios para determinar se uma página é uma duplicata de outra página na *internet* ou se é considerada canônica, isto é, a página que é selecionada para ser exibida nos resultados da pesquisa (Ribeiro *et al.*, 2020). A seleção da página canônica ocorre mediante a formação de grupos de páginas com conteúdo semelhante, das quais uma é eleita como a mais representativa do grupo. As demais páginas são consideradas versões alternativas que podem ser exibidas em diferentes contextos, tais como dispositivos móveis ou buscas específicas.

Para além do estabelecimento da página canônica, o *Google* coleta diversos sinais sobre essa página e seu conteúdo, os quais são utilizados em etapas subsequentes, quando ocorre a veiculação desta página nos resultados de pesquisa (Ribeiro *et al.*, 2020). Entre os sinais considerados estão o idioma da página, o país onde o conteúdo é localizado, a usabilidade da página, entre outros. Essas informações, coletadas e analisadas, são armazenadas no índice do *Google*, que consiste em um vasto banco de dados distribuído em milhares de computadores. Ressalta-se que a indexação não é um processo garantido, uma vez que nem todas as páginas processadas pelo *Google* são efetivamente indexadas.

Ademais, a indexação está intrinsecamente relacionada ao conteúdo da página e seus metadados. Problemas comuns que podem afetar a indexação incluem a baixa qualidade do conteúdo, restrições impostas pelas diretivas de robôs que impedem a indexação e dificuldades decorrentes do *design* do *site* que podem prejudicar a correta indexação.

A exibição dos resultados de pesquisa constitui uma etapa essencial do funcionamento dos motores de busca. Quando um usuário realiza uma consulta no *Google*, o motor de busca utiliza seus algoritmos para pesquisar o índice em busca de páginas que correspondam à consulta e, em seguida, retorna os resultados considerados de maior qualidade e relevância para o usuário. A determinação da relevância dos resultados leva em conta uma ampla gama de fatores, como a localização do usuário, o idioma utilizado e o dispositivo empregado (*desktop* ou telefone).

Vale mencionar que, além do *Google*, existem outras empresas que também atuam no mercado de motores de busca. Cada uma delas emprega algoritmos e estratégias próprias para fornecer resultados de pesquisa relevantes. Por exemplo, o *Bing*, motor de busca da

Microsoft, utiliza tecnologias como aprendizado de máquina e IA para aprimorar a precisão dos resultados

O *Bing*, motor de busca desenvolvido pela *Microsoft*, tem se destacado no mercado ao adotar tecnologias avançadas, como aprendizado de máquina e IA, com o objetivo de aprimorar a precisão dos resultados de pesquisa (Ribeiro *et al.*, 2020). Essas tecnologias, combinadas com a vasta quantidade de dados disponíveis na internet, permitem ao *Bing* oferecer resultados mais relevantes e personalizados para os usuários.

O aprendizado de máquina, uma subárea da IA, é essencial para o funcionamento eficiente do *Bing* (Ribeiro *et al.*, 2020). Por meio do treinamento de modelos algorítmicos com grandes conjuntos de dados, o motor de busca pode reconhecer padrões e estabelecer relações entre os diferentes elementos das consultas e dos documentos indexados (Ribeiro *et al.*, 2020). Isso possibilita que o *Bing* entenda o contexto e a intenção por trás das pesquisas dos usuários, resultando em resultados mais precisos e pertinentes.

A IA também desempenha um papel fundamental no *Bing*, especialmente na área de processamento de linguagem natural (Ribeiro *et al.*, 2020). Com algoritmos de processamento de linguagem natural avançados, o motor de busca é capaz de compreender as consultas dos usuários de forma mais precisa e identificar a correspondência entre as palavras-chave pesquisadas e os conteúdos disponíveis na *web* (Ribeiro *et al.*, 2020). Isso contribui para uma melhor correspondência entre as consultas dos usuários e os resultados apresentados, tornando a experiência de pesquisa mais satisfatória. Um estudo realizado por Ribeiro *et al.* (2020) destaca o uso do aprendizado de máquina no *Bing* para aprimorar a eficiência e a precisão dos resultados de pesquisa. Os autores enfatizam a importância do treinamento de modelos algorítmicos com grandes volumes de dados para melhorar a capacidade do motor de busca em identificar e atender às necessidades dos usuários.

Além disso, o *Baidu*, o principal motor de busca utilizado na China, também faz uso de tecnologias como aprendizado de máquina e IA para aprimorar sua precisão e relevância nos resultados de pesquisa (Zhang *et al.*, 2019). O *Baidu* tem se dedicado ao desenvolvimento de algoritmos personalizados, adaptados às características específicas do mercado chinês e à língua chinesa (Zhang *et al.*, 2019). Com isso, busca-se fornecer resultados mais relevantes e precisos para os usuários chineses, considerando as particularidades do idioma e as preferências culturais. Nesse sentido, uma pesquisa realizada por Zhang *et al.* (2019) abordou o uso de técnicas de aprendizado de máquina e a IA no *Baidu* para melhorar a precisão dos resultados de pesquisa em língua chinesa. Os autores ressaltam a importância de considerar as nuances e as peculiaridades do idioma chinês para o desenvolvimento de algoritmos eficientes e precisos.

Dessa forma, tanto o *Bing*, motor de busca da *Microsoft*, quanto o *Baidu*, principal motor de busca da China, utilizam tecnologias como aprendizado de máquina e IA para aprimorar a precisão dos resultados de pesquisa (Ribeiro *et al.*, 2020; Zhang *et al.*, 2019). Essas abordagens inovadoras permitem uma melhor compreensão das consultas dos usuários, a correspondência mais precisa com o conteúdo disponível na *web* e a entrega de resultados relevantes e personalizados.

Ao adotar o aprendizado de máquina, o *Bing* se beneficia do treinamento de modelos algorítmicos com grandes volumes de dados, o que resulta em uma maior capacidade de identificar e atender às necessidades dos usuários (Ribeiro *et al.*, 2020). Através do reconhecimento de padrões e do estabelecimento de relações entre elementos das consultas e dos documentos indexados, o *Bing* pode oferecer resultados mais precisos e pertinentes, aprimorando assim a eficiência de sua busca.

Além disso, a utilização de algoritmos de processamento de linguagem natural avançados pelo *Bing* permite uma compreensão mais precisa das consultas dos usuários e uma identificação mais eficiente da correspondência entre as palavras-chave pesquisadas e o conteúdo disponível na *web* (Ribeiro *et al.*, 2020). Isso contribui para uma experiência de pesquisa mais satisfatória, com resultados mais relevantes e precisos. No caso do *Baidu*, o motor de busca chinês, as tecnologias de aprendizado de máquina e IA também desempenham um papel fundamental na melhoria da precisão e relevância dos resultados de pesquisa em língua chinesa (Zhang *et al.*, 2019). Com algoritmos personalizados e adaptados às particularidades do mercado chinês e à língua chinesa, o *Baidu* busca fornecer resultados mais relevantes e precisos, levando em consideração as nuances e as preferências culturais do público chinês.

Os estudos de Ribeiro *et al.* (2020) e Zhang *et al.* (2019) evidenciam a importância do uso de técnicas de aprendizado de máquina e IA nos motores de busca, tanto para o *Bing* quanto para o *Baidu*, visando aprimorar a precisão dos resultados de pesquisa. Essas abordagens, aliadas à quantidade crescente de dados disponíveis e aos avanços tecnológicos, possibilitam que os motores de busca ofereçam uma experiência de pesquisa mais eficiente, com resultados mais relevantes e personalizados para os usuários.

3.3 A Busca Automatizada Utilizando a Linguagem *Python*

No início dos computadores, os programas não eram armazenados na máquina, em vez disso, a funcionalidade era programada para cada aplicativo. Um grande avanço ocorreu

com os programas armazenados, inicialmente escritos em linguagem de máquina. Tais linguagens são consideradas como a primeira geração de linguagens de programação. A segunda geração de linguagens de programação envolve linguagens de montagem, um pouco mais legíveis que as linguagens de máquina, mas ainda fortemente dependentes de uma arquitetura de conjunto de instruções particular. Finalmente, a terceira geração de linguagens de programação, cujo exemplo proeminente foi o *Fortran*, que apareceu entre as primeiras desta geração e ganhou enorme popularidade, forneceu uma abstração que separou o programador da arquitetura do conjunto de instruções da máquina, permitindo a portabilidade do código (Koffman, 1997).

Com código portátil, bibliotecas de *software* apareceram, acumulando conhecimento e experiência em programação, e assim, tornou-se uma atividade social. Bibliotecas de alto nível, como as de computação numérica, são hoje em dia muito ricas e completas, e é muito provável que um problema cotidiano que um programador enfrenta já esteja resolvido e incluído nela. Nessa atividade social, um programador conta com ferramentas de desenvolvimento de programas, como compiladores e ambientes de desenvolvimento integrados, desenvolvidos por outros programadores, bem como bibliotecas de *software*, se quiser programar com eficiência. Essa programação focada na solução de tarefas específicas, enquanto problemas gerais e frequentemente encontrados, já possui soluções de biblioteca prontamente disponíveis, levando a linguagens de “colagem”, projetada para fornecer inclusão eficiente de soluções de biblioteca e colá-las para resolver um problema específico. A evolução das bibliotecas de *software*, crescendo em tamanho e recursos diariamente, apoiou ainda mais esse conceito. Um exemplo proeminente de uma linguagem de programação que suporta o conceito de “*gluing*” é o *Python*. Projetado para ser legível, com sintaxe simples e clara, enquanto altamente extensível pela inclusão de bibliotecas de software, módulos nomeados, que podem ser usados confortavelmente usando um sistema de *namespace* conveniente (Pejovic, 2019).

A *Python* é uma linguagem de programação de código aberto e plataforma cruzada, que se tornou cada vez mais popular nos últimos dez anos. É uma linguagem de *script* de alto nível, interpretada, interativa e orientada a objetos, projetada para ser altamente legível, utilizando palavras-chave em inglês com frequência, enquanto outros idiomas usam pontuação e têm menos construções sintáticas do que outros idiomas (Menezes, 2014).

Além disso, a *Python* é uma linguagem de programação de propósito geral de alto nível; como o código é compilado automaticamente em código de *bytes* e executado, o *Python* é adequado para uso como linguagem de *script*, linguagem de implementação de aplicativos da

web etc. Por ser estendido em C e C++², o *Python* pode fornecer a velocidade necessária até mesmo para tarefas de computação intensiva. Por causa de suas fortes construções de estruturação, blocos de código aninhados, funções, classes, módulos e pacotes, e seu uso consistente de objetos e programação orientada a objetos, o *Python* permite escrever aplicativos lógicos e claros para tarefas pequenas e grandes (Kuhlman, 2012).

Segundo Kuhlman (2012), as principais características do *Python* são:

- Tipos de dados de alto nível embutidos: *strings*, listas, dicionários etc;
- As estruturas de controle usuais: *if*, *if-else*, *if-elif-else*, *while*, mais um poderoso iterador de coleção (*for*);
- Múltiplos níveis de estrutura organizacional: funções, classes, módulos e pacotes. Estes ajudam na organização do código;
- Um excelente e amplo exemplo é a biblioteca padrão do *Python*. Compilar dinamicamente para código de *byte*. O código-fonte é compilado para código de *byte* sem uma etapa de compilação separada;
- Os módulos de código-fonte também podem ser "pré-compilados" para arquivos de código de *bytes*;
- Orientado a objetos - *Python* fornece uma maneira consistente de usar objetos: tudo é um objeto;
- E, em *Python*, é fácil implementar novos tipos de objetos, chamados de classes na programação orientada a objetos;
- Extensões em C e C++ - Módulos de extensão e tipos de extensão podem ser escritos à mão.

Python usa tipagem dinâmica e uma combinação de contagem de referência e um coletor de lixo de detecção de ciclo para gerenciamento de memória. Usa resolução dinâmica de nomes, que vincula nomes de métodos e variáveis durante a execução do programa. Seu *design* oferece algum suporte para programação funcional na tradição *Lisp*. Possui funções de filtro, mapa e redução; compreensões de lista, dicionários, conjuntos e expressões geradoras. É

² C e C++ são linguagens de programação amplamente utilizadas no desenvolvimento de *software*. C é uma linguagem de programação de alto nível criada por Dennis Ritchie em 1972. Enquanto C++ é uma extensão da linguagem C, desenvolvida por Bjarne Stroustrup nos anos 1980. Ele adicionou recursos de programação orientada a objetos ao C, tornando-o uma linguagem mais poderosa e versátil. O C++ é amplamente utilizado em desenvolvimento de jogos, aplicativos de alto desempenho e sistemas complexos. Disponível em: <https://blog.betrybe.com/linguagem-de-programacao/cpp/>. Acesso em: 6 out. 2023.

uma linguagem de programação multiparadigma. Programação orientada a objetos e programação estruturada são totalmente suportadas, e muitos de seus recursos suportam programação funcional e programação orientada a aspectos (Ramalho, 2022).

Ao invés de construir toda a sua funcionalidade em seu núcleo, o *Python* foi projetado para ser altamente extensível por meio de módulos. Essa modularidade compacta o tornou particularmente popular como meio de adicionar interfaces programáveis a aplicativos existentes. A visão de Van Rossum de uma pequena linguagem central com uma grande biblioteca padrão e um interpretador facilmente extensível surgiu de suas frustrações com o ABC, que defendia a abordagem oposta (Ramalho, 2022).

A pesquisa na *web* pode ser automatizada usando o *script Python*. Isso pode ser feito usando o *Selenium*, uma ferramenta de automação do navegador. O *Selenium* é um *framework* portátil para testar aplicações *web*. Ele pode executar automaticamente as mesmas interações que precisa executar manualmente. A automação faz três coisas principais: substitui tarefas humanas tediosas, minimiza erros e usa ajuda de caso. Empresas a caminho da transformação digital.

3.4 Design Science Research

Os métodos fundamentados na DSR buscam não apenas compreender a organização em sua forma atual, mas também desenvolver conhecimentos que possam aprimorá-la. Essa abordagem tem como premissa central a redução do distanciamento entre a academia e as organizações, bem como entre a teoria e a prática, como se evidencia na perspectiva dos renomados autores.

Além disso, as pesquisas orientadas pela DSR frequentemente possuem objetivos distintos e geram conhecimentos diferenciados. Em muitos casos, uma pesquisa realizada no contexto das áreas mencionadas anteriormente, como engenharia e gestão CI, entre outros, não se limita à exploração, descrição ou explicação do problema, mas também se dedica a desenvolver propostas para sua solução. Nesse sentido, o resultado esperado de uma pesquisa pode ser a prescrição de uma solução ou a concepção de um produto. Esses objetivos não podem ser alcançados por meio dos métodos de pesquisa fundamentados exclusivamente nas ciências tradicionais (Dresch *et al.*, 2015, p. 124).

No contexto das pesquisas baseadas em DSR, há discussões relevantes entre os autores acerca dos objetivos e do conhecimento gerado por essa abordagem. Segundo Peffers *et al.* (2020), a DSR difere dos métodos de pesquisa tradicionais ao não se limitar apenas a

explorar, descrever ou explicar um problema, mas também ao buscar desenvolver soluções práticas e aplicadas para esse problema.

Essa perspectiva orientada para a ação e para a produção do artefato permite uma contribuição direta para a prática, fornecendo soluções concretas e aplicáveis.

O Quadro 1 apresenta uma compilação das principais definições que serão abordadas neste capítulo, proporcionando uma visão resumida e abrangente desses conceitos fundamentais.

Quadro 1 – Conceitos e Descrições da Metodologia DSR.

Conceito	Descrição
<i>Design Science</i>	É uma abordagem científica que busca consolidar conhecimentos e desenvolver soluções para melhorar sistemas existentes, resolver problemas e criar artefatos.
Artefato	É uma criação humana que serve como interface entre o ambiente interno e externo de um sistema específico.
Soluções Satisfatórias	Soluções suficientemente adequadas para o contexto em questão. As soluções devem ser viáveis, não necessariamente ótimas.
Classes de problemas	Organização que orienta a trajetória e o desenvolvimento do conhecimento no âmbito da Design Science e são suficientemente adequadas para o contexto em questão. Elas podem não ser ótimas, mas devem ser viáveis e atender às necessidades específicas.
Validade pragmática	Visa garantir a utilidade da solução proposta para o problema, levando em consideração o custo/benefício, as particularidades do ambiente de aplicação e as necessidades reais dos interessados na solução.

Fonte: Elaborado por Ocante António Ié (2023) baseado em Dresch *et al.* (2015, p. 59)

Nesse sentido, Dresch *et al.* (2015) destacam que a DSR proporciona uma oportunidade para que os pesquisadores possam transcender a mera investigação teórica, promovendo a transferência de conhecimento entre a academia e as organizações. Ao desenvolver soluções práticas, a pesquisa baseada em DSR busca atender às necessidades reais das academias na produção do conhecimento e em organizações na contribuição das melhorias dos processos e práticas existentes. Essa abordagem alinha-se com a demanda atual por pesquisas aplicadas e com impacto prático, resultando em um conhecimento que vai além da mera compreensão conceitual.

Nesse contexto, a pesquisa baseada em evidências e conhecimento faz essa escolha com base na síntese dos resultados de uma revisão sistemática da literatura sobre o problema em questão. Além disso, a DSR uma forma de produção de conhecimento científico que busca resolver problemas do mundo real e fazer contribuições científicas de natureza prescritiva. Essa

abordagem se mostra especialmente adequada para pesquisas nas áreas de engenharia e gestão. É importante ressaltar que a pesquisa explicativa não é excluída pela DSR, pois pode ser combinada com a pesquisa prescritiva. A *Design Science* valoriza a construção do conhecimento em função da ação, envolvendo o estudo do projeto e a criação de artefatos.

Os estudos contemporâneos sobre as principais diferenças entre a ciência tradicional e a DSR têm ampliado as discussões nessa área. Autores como Hevner *et al.* (2004) enfatizam que a ciência tradicional busca entender e explicar os fenômenos organizacionais por meio de uma abordagem descritiva e analítica. Por outro lado, a DSR se concentra no projeto e na produção de sistemas que ainda não existem, com o objetivo de resolver problemas e alcançar melhores resultados. Essa diferença é evidenciada na natureza dos objetos de estudo, em que a ciência tradicional trata os fenômenos organizacionais como objetos empíricos, enquanto a DSR considera as questões organizacionais e sistemas como objetos artificiais, exigindo intervenções não rotineiras.

Além disso, os autores contemporâneos têm ressaltado a visão do conhecimento na ciência tradicional e na DSR. Segundo Pries-Heje *et al.* (2005), a ciência tradicional busca um conhecimento representacional, em que o pensamento é descritivo e analítico, buscando formulações de hipóteses e testes. Já a DSR adota uma abordagem pragmática, em que o conhecimento está a serviço da ação, sendo normativo e sintético. Isso significa que DSR reconhece a unicidade de cada situação e busca soluções ideais, baseando-se em informações limitadas, participação, discurso e experimentação pragmática.

Outra discussão relevante nos estudos atuais é o foco no desenvolvimento da teoria. Gregor e Hevner (2013) destacam que a ciência tradicional se concentra na descoberta de relações causais gerais entre variáveis por meio de hipóteses e testes. Por outro lado, a *Design Science* tende a se mover além dos limites iniciais da situação, buscando o projeto e o desenvolvimento de novos artefatos. Essa abordagem vai além da análise das relações causais, envolvendo intervenções não rotineiras e a criação de sistemas ideais.

O Quadro 2 mostra as principais diferenças entre a ciência tradicional e a *Design Science*:

Quadro 2 - Diferenças entre a Ciência Tradicional e a *Design Science*.

Elemento	Ciência Tradicional	<i>Design Science</i>
Propósito.	Entender fenômenos organizacionais por meio de objetividade consensual.	Projetar e produzir sistemas novos e modificar situações existentes para solucionar problemas.

Objetivo da pesquisa.	Explorar, descrever, explicar e prever.	Prescrever soluções para problemas.
Visão do conhecimento.	Representacional, busca por conhecimentos gerais e válidos.	Pragmática, conhecimento a serviço da ação.
Natureza dos objetos.	Fenômenos organizacionais como objetos empíricos.	Questões organizacionais e sistemas como objetos artificiais com propriedades mal definidas.
Foco no desenvolvimento	Descoberta de relações causais gerais entre variáveis.	Projeto e desenvolvimento de novos artefatos que vão além das definições iniciais da situação.

Fonte: Elaborado por Ocante Antônio Ié (2023), baseado em Dresch *et al.* (2015, p. 61)

Para efetivamente impulsionar a ação, é imprescindível adentrar nas distintas perspectivas do conhecimento e compreender sua relação intrínseca com a prática. Nesse sentido, Gibbons *et al.* (1994) destacam a existência de dois modos de produção de conhecimento: o tipo I, que se restringe ao âmbito acadêmico e está confinado a uma única disciplina, e o tipo II, que se apresenta como transdisciplinar e visa solucionar problemas que emergem no contexto de aplicação.

O modo de produção de conhecimento tipo I caracteriza-se pela abordagem disciplinar, em que a geração de conhecimento se dá dentro dos limites restritos de uma determinada área de estudo. Essa perspectiva de produção de conhecimento geralmente concentra-se em questões teóricas e na exploração de novas percepções dentro do campo delimitado. Entretanto, seu alcance pode se mostrar limitado ao enfrentar problemas complexos e interdisciplinares que emergem na prática.

Por outro lado, o modo de produção de conhecimento tipo II busca transpor as limitações da abordagem disciplinar, adotando uma postura transdisciplinar. Esse modo de produção de conhecimento envolve a colaboração e a integração de múltiplas disciplinas para abordar os problemas práticos do mundo real. Reconhece-se que os desafios enfrentados na prática frequentemente exigem uma compreensão holística e uma visão abrangente que ultrapasse as fronteiras de uma única disciplina.

A pesquisa em DSR, de acordo com Dresch *et al.* (2015), privilegia o conhecimento tipo II como o mais adequado, embora reconheça a relevância do conhecimento tipo I como suporte teórico para a pesquisa. Nesse sentido, a abordagem transdisciplinar e construtivista do conhecimento tipo II, conforme proposto por Gibbons *et al.* (1994), rompe com a concepção linear de transferência de conhecimento, buscando uma interação dinâmica e colaborativa entre diferentes disciplinas.

A pesquisa tradicional, predominantemente associada ao conhecimento tipo I, tem sido caracterizada pelo emprego de métodos indutivos e dedutivos. Esses métodos têm sido amplamente discutidos na literatura, ressaltando suas abordagens e aplicações. No entanto, a DSR traz à tona um terceiro método, denominado “Abdução”. Segundo Dresch *et al.* (2015), a abdução é reconhecida como um processo altamente criativo, sendo especialmente indicada para a compreensão de situações-problema complexas, devido à sua natureza intrinsecamente reflexiva e criativa.

Segue a síntese o Quadro 3 das abordagens mencionadas, para uma melhor compreensão.

Quadro 3 – Os Componentes, Pesquisa Tradicional e DSR.

Componente	Pesquisa Tradicional	<i>Design Science</i>
Conhecimento	Tipo I	Tipo II
Métodos	Indutivos e dedutivos	Abdução
Características	Ênfase em abordagens indutivas e dedutivas.	Ênfase na abordagem abdutiva.
Aplicações	Ampla discussão na literatura	Abordagem emergente, com maior destaque na <i>Design Science Research</i>
Natureza	Racional e analítica	Reflexiva e criativa
Situações-problema	Adequadas para problemas com relação causa-efeito bem definida	Indicada para situações-problema complexas e menos estruturadas

Fonte: Elaborado por Ocante Antônio Ié (2023), baseado em Dresch *et al.* (2015).

Nesse contexto, a pesquisa em *DSR* se destaca por sua ênfase na busca por soluções inovadoras e aplicáveis, baseadas em uma abordagem transdisciplinar e na utilização do raciocínio Abdutivo. Essa abordagem se alinha com as demandas do mundo contemporâneo, que exige respostas efetivas e integradas para os desafios complexos enfrentados nas áreas de engenharia, gestão, entre outros campos. Ao adotar o conhecimento tipo II e explorar o potencial criativo da Abdução, a pesquisa em *Design Science* se posiciona como uma abordagem promissora para a produção de conhecimento científico que busca aprimorar sistemas existentes, resolver problemas e criar artefatos.

Uma das características distintivas da *DSR* é a sua abertura para a utilização de múltiplos métodos de pesquisa. Dresch *et al.* (2015) ressalta que o método abdutivo desempenha um papel crucial quando o pesquisador está propondo soluções inovadoras para

resolver problemas complexos, exigindo um pensamento criativo e atividades exploratórias significativas. Por outro lado, em situações em que é necessário realizar uma avaliação lógica dos aspectos relacionados aos artefatos, o Método Dedutivo se apresenta como a abordagem mais indicada.

Embora a DSR seja considerada uma metodologia robusta, rigorosa e completa, existem casos em que a incorporação de abordagens complementares pode ser mais conveniente. Nessas circunstâncias, métodos tradicionais, como estudo de caso e pesquisa-ação, podem ser utilizados para enriquecer a pesquisa.

O estudo de caso, por exemplo, desempenha um papel importante ao produzir conhecimento teórico-analítico e formalizar artefatos eficazes, que possam ser úteis para outros fins, contribuindo assim para a busca de generalizações, um aspecto enfatizado pela DSR (Van Aken, 2004).

3.5 Aplicação de Diretrizes de *Design Science Research*

À medida que a pesquisa avança, é evidente que a relevância do problema é endossada por cada etapa da metodologia DSR. A abordagem da pesquisa qualitativa espelha o foco na eficácia proporcionada pelos algoritmos de busca, tal como destacado na citação de Jansen *et al.* (2008). A metodologia DSR, citada no método da pesquisa, fornece o arcabouço para desenvolver artefatos como o Artefato descrito – um protótipo de algoritmos de motores de busca.

Para assegurar o rigor da pesquisa, a metodologia emprega técnicas de coleta e tratamento de dados específicos, correspondendo às categorias de técnica de coleta de dados e técnica de tratamento de dados da tabela anterior. Essas técnicas permitem que os dados sejam analisados e interpretados de maneira sistemática, garantindo a validade e a qualidade da pesquisa.

A avaliação realizada, que envolve testes e análise de utilização, complementa o processo de desenvolvimento e pesquisa, assegurando que os artefatos criados sejam eficazes e refinados. A contribuição da pesquisa encaixa-se de forma coerente, uma vez que a metodologia DSR visa à geração de conhecimento prático e teórico para avançar a compreensão dos algoritmos de busca na perspectiva da CI.

Finalmente, a comunicação da pesquisa se alinha ao propósito de compartilhar os resultados e as descobertas com a comunidade científica e profissional da área de CI. Portanto, os dois quadros estão intrinsecamente interconectados, pois cada categoria na metodologia DSR

contribui de maneira sequencial e lógica para a execução eficaz da pesquisa, do entendimento à aplicação e à disseminação (Quadro 4).

Quadro 4 - Aplicação da Diretriz.

Diretriz	Aplicação da Diretriz
Relevância do problema	Para as pessoas sem compreensão do funcionamento dos motores de busca, a tecnologia torna as coisas mais fáceis. Para os usuários, a tecnologia torna as coisas possíveis. A presente pesquisa tem como foco a facilidade provida por algoritmos de motores de busca para aumentar a eficácia na compreensão de mecanismos de pesquisa de diferentes fins.
Artefato	O artefato desenvolvido consiste em um protótipo de algoritmos de motores de busca.
Rigor da pesquisa	Para cada fase do desenvolvimento da pesquisa, foram utilizados um ou mais conceitos para garantir o rigor do estudo. A avaliação das abordagens de diferentes literaturas, análise da utilização dos componentes linguísticos, bibliotecas e frameworks foram adotados para assegurar a qualidade da pesquisa durante as fases de desenvolvimento.
Avaliação	A avaliação de usabilidade envolveu testes utilizando técnicas de observação de uso e análise da utilização das ferramentas.
Contribuição da pesquisa	A pesquisa contribui para avançar o conhecimento no campo da CI, especificamente no que diz respeito à compreensão e aprimoramento dos algoritmos de automação dos motores de busca.

Fonte: Elaborado por Ocante Antônio Ié (2023), baseado em Hevner *et al.* (2004).

3.5.1 A Importância do Problema

No contexto desta pesquisa, a compreensão do funcionamento dos algoritmos de motores de busca é de suma relevância. A literatura destaca a importância de adotar uma abordagem centrada no usuário ao estudar esses algoritmos (Jansen *et al.*, 2008). Ao compreender como os usuários interagem com os motores de busca, podemos identificar limitações e melhorar a experiência de busca.

É relevante abordar o problema enfrentado pelos usuários na busca por informações. A compreensão dos algoritmos de motores de busca permite identificar possíveis falhas no processo de busca e sugerir melhorias para aumentar a eficácia e a precisão dos resultados de pesquisa (Jansen *et al.*, 2008). Ao entender as nuances do funcionamento dos algoritmos, podemos desenvolver estratégias para otimizar a RI e fornecer resultados mais relevantes aos usuários.

Nesse sentido, o estudo destaca a importância de considerar as necessidades e expectativas dos usuários ao projetar algoritmos de motores de busca. Uma abordagem centrada no usuário envolve a análise do comportamento de busca, as preferências de interação e as necessidades informacionais dos usuários. Compreender esses aspectos é fundamental para desenvolver algoritmos que atendam de forma efetiva às demandas dos usuários.

Portanto, a compreensão dos algoritmos de motores de busca e sua aplicação na perspectiva do usuário é essencial para melhorar a experiência de busca e obter resultados mais relevantes. Ao adotar uma abordagem centrada no usuário, podemos identificar oportunidades de aprimoramento e desenvolver algoritmos que atendam às necessidades e expectativas.

3.5.2 Avaliação de Artefatos

No desenvolvimento e avaliação de artefatos nesta pesquisa, foi criado e implementado um exemplo de *Web Crawler* utilizando a linguagem de programação *Python*. Esse artefato permite a realização de buscas automatizadas e explora a interação entre o processo de automação e os algoritmos de busca.

A abordagem da DSR valoriza a criação e avaliação de artefatos como uma forma de contribuir para o avanço do conhecimento teórico-prático. Conforme destacado por Bax (2015), a DSR permite ao pesquisador criar artefatos com base em conjecturas teóricas, proporcionando a geração de novos dados que contribuem para a evolução do conhecimento existente.

Ao desenvolver o *Web Crawler* como artefato, busca-se fornecer uma solução prática para os desafios identificados na pesquisa. Nesse sentido, permite explorar o comportamento dos algoritmos dos motores de busca e realizar buscas automatizadas de informações na *web*. Dessa forma, é possível obter dados relevantes sobre os processos de automação e sua interação com os algoritmos de busca.

A avaliação desse artefato é fundamental para verificar sua eficácia e contribuição para o campo da CI. Através da avaliação, novos dados são gerados e os esclarecimentos são obtidos, permitindo aprimorar o conhecimento teórico existente. Essa avaliação pode ser realizada por meio de testes, experimentos e comparações com outros artefatos ou abordagens existentes.

A escolha de desenvolver e avaliar artefatos está em consonância com a abordagem da DSR, que valoriza a aplicação prática e a criação de soluções para problemas identificados. Ao utilizar a metodologia DSR e desenvolver um artefato como parte da pesquisa, busca-se contribuir para o avanço da CI e para a otimização dos processos de busca de informações na era digital.

A aplicação da metodologia de pesquisa no contexto desta investigação visa compreender o comportamento dos algoritmos nos processos de automação dos motores de busca, sob a perspectiva da CI. Isso proporcionará uma compreensão dos limites de

comunicação e da busca na *web* pelos usuários. Além disso, serve para guiar o desenvolvimento e adaptá-la.

3.5.3 Rigor da Pesquisa

No âmbito deste contexto, a pesquisa se destaca como uma faceta de significativa envergadura, delineando a profundidade e a meticulosidade que permeiam o processo investigativo. A noção de rigor, entremeada pela sua conotação de precisão, exatidão e integridade, infunde-se na essência da pesquisa, configurando-se como um alicerce indelével que sustenta a credibilidade e a validade intrínseca do estudo.

A pesquisa emerge como uma expressão do compromisso inabalável com os preceitos científicos e acadêmicos, transcende meros exercícios de superficialidade. Nesse contexto, a abordagem adotada para avaliar a pesquisa é de singular importância. Sob o prisma de um exame meticuloso, esta investigação denota uma adesão resoluta aos cânones estabelecidos pelos mestres da CI e da Gestão de Informação.

A busca de conhecimento e a inquisição diligente são o cerne dessa empreitada. A interação contínua com *stakeholders*, usuários finais e especialistas proeminentes, alinhada à sua perspicácia, demonstra um comprometimento intrínseco em compreender a complexidade das necessidades informacionais. Nesse contexto, as reflexões de Marchand e Davenport (2000) emergem como faróis orientadores, delineando a necessidade de uma interação colaborativa, que transcende meros formalismos, para capturar a sutileza subjacente aos requisitos e metas.

Aprofundando-se nessa jornada de pesquisa, é inequívoco que o rigor não é uma mera formalidade. O algoritmo de aplicação *web* de *web scraping* se desvela como um microcosmo intricado, onde a identificação de fontes de dados relevantes, o tratamento cuidadoso das informações coletadas e a engenhosidade na apresentação dos resultados coalescem em uma sinfonia harmoniosa. Tal processo, que remanesce sob o escrutínio de Batista (2014), emerge como um testemunho do compromisso com o rigor, onde a gestão da informação é evocada como um mestre de cerimônias que orchestra a coleta, organização e disseminação dos dados.

Em consonância, os testes exaustivos concebidos para o projeto delineiam uma etapa crítica na aferição da pesquisa. As nuances dos cenários de teste, os casos meticulosamente delineados e as considerações de segurança ilustram uma abordagem que não se resigna a falhas, que não cede ao superficial, mas que abraça o desafio de um escrutínio minucioso. A deliberação do impacto real das soluções implementadas, o confronto entre

resultados esperados e reais, bem como a empatia intrínseca ao teste de usuário final, denotam uma crença inalterada na importância de uma pesquisa que transcende as páginas dos manuscritos.

Portanto, o rigor desta pesquisa ressurgiu como um farol orientador que ilumina cada etapa, que permeia cada linha de código e que respira vida nas interfaces do aplicativo. Ao ressoar na harmonia dos autores proeminentes, a pesquisa reitera a sua devoção à precisão e à integridade, ao entregar uma aplicação que não é meramente um algoritmo, mas uma destilação refinada do conhecimento e do esforço empírico, forjados sob a égide do rigor.

No contexto deste trabalho, se vale do embasamento teórico fornecido por eminentes estudiosos na área de CI e Gestão de Informação. A imersão deste conhecimento enriquece ainda mais a compreensão das implicações e da relevância no processo de pesquisa e desenvolvimento.

Nas palavras de Riemenschneider *et al.* (2002), o conceito de rigor está intrinsecamente associado à investigação científica, representando a consistência metodológica e a aderência a princípios de pesquisa rigorosos. A pesquisa transcende a mera exploração casual, adentrando a esfera de um exame criterioso que envolve a formulação precisa de perguntas, a aplicação de métodos sólidos e a validação metódica dos resultados.

No domínio da CI, a essência do rigor encontra eco nas reflexões de Batista (2014), cuja obra "Gestão da Informação: Como transformar informações em ativos para a empresa" serve como um farol que guia o entendimento da gestão informacional como um processo estratégico. A intersecção entre a gestão de informações e o processo de pesquisa é notável, uma vez que a aplicação *web* de *web scraping* converge a busca por informações relevantes e a sua transformação em ativos tangíveis.

A conceituação do rigor, sob a égide da Gestão de Informação, se alinha com a noção de captura, organização e armazenamento de dados, conforme apontado por Batista (2014). A análise minuciosa dos dados coletados, a sua organização lógica, a sua subsequente apresentação revela o compromisso com a fidedignidade da informação e a sua utilidade para a tomada de decisões.

A abordagem proposta para avaliar o rigor desta pesquisa, notadamente no que tange aos testes e validações, ecoa as diretrizes delineadas por Marchand e Davenport (2000). O engajamento colaborativo com *stakeholders* e usuários finais para compreender as necessidades e expectativas configura uma estratégia alinhada com a busca de soluções que sejam verdadeiramente úteis, a exemplo da aplicação de *web scraping* que busca extrair informações relevantes.

Além disso, a incorporação de métodos de teste de segurança e desempenho, bem como a constante avaliação dos resultados em relação às expectativas, demonstra a consideração criteriosa do rigor metodológico. Sob esse prisma, a pesquisa encara de frente a demanda por veracidade e solidez, conforme preconizado por Riemenschneider *et al.* (2002).

Portanto, as contribuições teóricas de autores como Batista (2014), Marchand e Davenport (2000), e Riemenschneider *et al.* (2002) entre outros aqui debatidos, oferecem um arcabouço sólido e coerente para a apreciação do rigor nesta pesquisa. A interseção entre a CI e a Gestão de Informação com a aplicação de *web scraping* se desvela como uma simbiose que reflete a busca incessante pela precisão, integridade e relevância informacional. Ao embasar-se em preceitos robustos, essa pesquisa enriquece o panorama do rigor científico e potencializa a confiança nas soluções concebidas.

3.5.4 Avaliação da Pesquisa

A avaliação da pesquisa é uma atividade essencial no contexto científico, representando o processo crítico e rigoroso pelo qual se busca avaliar a qualidade, relevância e contribuição de um estudo específico para o corpo de conhecimento existente. Este ato avaliativo transcende a mera apreciação, adentrando no domínio da análise criteriosa e sistemática, alicerçada na reflexão profunda sobre os princípios teóricos, a metodologia empregada e os resultados alcançados, à luz do panorama histórico e conceitual delineado na sequência textual ora percorrida.

A análise da pesquisa em foco, delineada a partir do exame dos aludidos textos anexados, revela-se como uma abordagem de inegável pertinência no âmbito da CI. A articulação de conceitos que evoluíram a partir dos alvares do século XVI, com o advento das sociedades científicas, culminando no surgimento da CI como campo multidisciplinar, proporciona um enquadramento sólido e contextualmente embasado para a compreensão da pesquisa em avaliação.

A exploração da emergência da CI em resposta à "explosão da informação", fomentada pelo avanço tecnológico e pela Segunda Guerra Mundial, é respaldada por uma abordagem historiográfica sólida. A descrição das influências de figuras como Paul Otlet, Norbert Wiener, Claude Shannon e Warren Weaver enriquece o panorama, evidenciando a gênese teórica e a adoção de perspectivas interdisciplinares que consolidaram a disciplina.

A interligação entre a CI e a pesquisa de recuperação da informação é especialmente saliente, refletindo a essência da preocupação primordial com a organização, disseminação e

utilização do conhecimento em face do constante crescimento informacional. A abordagem da evolução dos sistemas, das técnicas e das tecnologias empregadas para a recuperação da informação delineia um percurso consistente, respaldado em marcos históricos e tecnológicos relevantes, culminando na delimitação da pertinência da pesquisa realizada.

Com relação à metodologia adotada, é possível identificar uma estrutura coerente e sistematizada. A análise parte da apresentação dos principais conceitos da CI, traçando sua evolução histórica até a contemporaneidade. A exploração da relevância dos algoritmos de busca na recuperação da informação é realizada de maneira analítica e detalhada, abordando suas características, funcionamento e implicações éticas e de transparência. A conexão entre a teoria e a prática é demonstrada por meio da exploração de modelos de recuperação da informação, o que atesta o rigor na exposição do conteúdo.

Contudo, para uma avaliação mais aprofundada, foi possível a explicitação da metodologia de pesquisa utilizada na elaboração deste estudo, como a seleção dos textos, a análise crítica das fontes e a fundamentação dos argumentos. Além disso, uma abordagem mais aprofundada das implicações contemporâneas da recuperação da informação, à luz das preocupações com privacidade, com viés algorítmico e controle da informação, tomaram parte da abordagem nesta pesquisa, para enriquecer ainda mais a análise.

Em síntese, a pesquisa em avaliação destaca-se por sua abordagem sólida e bem embasada, que perpassa por marcos históricos, teorias fundamentais e a evolução das práticas de recuperação da informação. A articulação entre o desenvolvimento histórico e a contemporaneidade da CI é hábil, proporcionando uma visão panorâmica da pesquisa. Apesar da abordagem consistente, os aprofundamentos na metodologia e a exploração mais profunda das implicações atuais são contribuintes para uma análise ainda mais abrangente e enriquecedora.

3.5.5 Comunicação da Pesquisa.

A pesquisa científica representa um veículo fundamental para a evolução do conhecimento em diversos campos, e na área da CI, essa prática não é diferente. Uma parte essencial desse processo é a comunicação dos resultados e descobertas obtidas, permitindo que a comunidade científica e profissional possa acessar, compreender e avaliar o valor das investigações realizadas. No contexto específico desta pesquisa, que se concentra na compreensão dos motores de busca e na automação de processos de busca, a comunicação

assume um papel ainda mais relevante, uma vez que aborda aspectos centrais da recuperação da informação, tecnologias de algoritmos e seu impacto na sociedade contemporânea.

Para efetivamente comunicar os achados desta pesquisa, será adotada uma abordagem baseada nas diretrizes de DSR, uma metodologia que busca a criação de soluções inovadoras por meio da integração entre pesquisa científica e prática aplicada. A DSR sugere a criação e avaliação de artefatos, como modelos, métodos, processos ou sistemas, que se alinham com os objetivos da pesquisa. Neste contexto, a pesquisa sobre algoritmos de motores de busca pode ser enquadrada como uma investigação orientada para a criação de conhecimento aplicado.

A estrutura da comunicação da pesquisa seguirá os princípios da DSR, visando apresentar de forma clara e estruturada os resultados obtidos e sua relevância para a comunidade acadêmica e profissional da CI. Os principais elementos que compõem a abordagem de comunicação, na pesquisa identificou a necessidade de compreender os algoritmos de motores de busca e sua aplicação na recuperação da informação. A partir dessa identificação, foram propostas soluções na forma de análises teóricas, revisões de literatura e investigações práticas.

Com a identificação precisa do problema, sendo o ponto de partida para qualquer pesquisa científica significativa, originou-se explosão de informações digitais e a dependência crescente das pessoas em motores de busca para acessar e organizar esses dados, levantando a questão fundamental: como esses algoritmos funcionam e como eles afetam o acesso à informação? Essa problemática foi abordada por meio da proposição de soluções, que incluem análises teóricas, revisões de literatura e investigações práticas, visando preencher a lacuna de conhecimento e proporcionar compreensão para a comunidade acadêmica e profissional da CI.

Os artefatos desenvolvidos incluem uma fundamentação teórica abrangente sobre CI, algoritmos de motores de busca e recuperação de informação. Essa fundamentação serviu de base para a compreensão dos conceitos chave e discussões relevantes no campo.

O estudo empregou uma metodologia de revisão de literatura, explorando as bases teóricas que sustentam a compreensão dos algoritmos e sua aplicação na recuperação da informação. Foram analisados modelos de recuperação da informação, as implicações éticas e os desafios relacionados à transparência e eficácia dos algoritmos de busca. No âmbito da pesquisa científica, os artefatos assumem a forma de modelos teóricos, metodologias e resultados concretos que contribuem para a compreensão e resolução do problema em questão. Os artefatos desenvolvidos neste estudo englobam uma base teórica sólida que abrange desde a evolução da CI até os conceitos-chave dos algoritmos de motores de busca.

A criação desses artefatos demonstra um comprometimento com a construção de conhecimento relevante e útil para a comunidade, indo além da mera análise teórica. A metodologia adotada reflete a abordagem empregada para responder às questões de pesquisa e alcançar os objetivos propostos. Neste estudo, a revisão de literatura foi escolhida como a abordagem principal, permitindo a exploração das bases teóricas que sustentam a compreensão dos algoritmos e sua aplicação na recuperação da informação. Essa metodologia permitiu uma análise crítica e sistemática dos modelos de recuperação da informação, da transparência e dos desafios éticos associados aos algoritmos de busca, contribuindo para uma compreensão abrangente e aprofundada do tema.

A pesquisa proporcionou uma visão aprofundada dos algoritmos de motores de busca, sua relevância na recuperação da informação e os desafios éticos e técnicos associados. Esses resultados contribuem para a compreensão da CI e seu papel na sociedade contemporânea. Resultou em uma série de contribuições significativas para a área da CI, assim, os resultados obtidos proporcionam uma visão aprofundada dos algoritmos de motores de busca, sua relevância na recuperação da informação e os desafios éticos e técnicos inerentes. A construção dessa base de conhecimento sólida é fundamental para informar debates e discussões na comunidade acadêmica e profissional, permitindo a formulação de estratégias mais informadas para lidar com a crescente complexidade dos sistemas de busca.

A abordagem foi avaliada à luz das diretrizes de DSR, destacando sua contribuição para o conhecimento científico na área de CI. A análise da relevância dos algoritmos de busca na gestão da informação e na sociedade ressalta a importância dessa pesquisa para a comunidade acadêmica e além. Os *insights* obtidos permitem aos pesquisadores e profissionais da área adotar uma abordagem mais crítica e fundamentada no desenvolvimento e aprimoramento de sistemas de recuperação da informação.

Os resultados conquistados podem ser utilizados por profissionais da área de CI para aprimorar a eficácia e transparência dos sistemas de busca. Além disso, a conscientização sobre os desafios éticos envolvidos na classificação e ordenação dos resultados de busca pode orientar a tomada de decisões responsáveis.

Os resultados desta pesquisa têm implicações práticas significativas. Profissionais da área de CI podem utilizar a percepção obtida para melhorar a eficácia e transparência dos sistemas de busca, garantindo que os usuários tenham acesso a informações relevantes e confiáveis. A conscientização sobre os desafios éticos relacionados à classificação e ordenação dos resultados de busca também pode orientar a tomada de decisões responsáveis por parte dos desenvolvedores e administradores desses sistemas.

A comunicação da pesquisa não se encerra aqui. Ela motiva futuras investigações sobre a melhoria dos algoritmos de busca, a transparência algorítmica e a minimização de vieses, enriquecendo o debate acadêmico e prático. A pesquisa também pode servir como base para o desenvolvimento de ferramentas educacionais que aumentem a compreensão pública sobre o funcionamento dos motores de busca, empoderando os usuários a tomar decisões informadas sobre suas interações *on-line*.

A comunicação da pesquisa científica neste contexto não apenas compila informações e resultados, mas também direciona a discussão e o desenvolvimento alinhado com as diretrizes de DSR. Essa abordagem reforça a conexão entre a pesquisa e a prática, demonstrando seu valor não apenas como uma exploração teórica, mas também como um esforço para contribuir para a evolução da CI e para a sociedade em geral.

Portanto, pesquisa sobre algoritmos de motores de busca, através de uma abordagem rigorosa e sistemática, oferece esclarecimento sobre o funcionamento dos algoritmos, seu impacto na sociedade e desafios relacionados, estabelecendo as bases para discussões informadas e aprimoramentos contínuos nos sistemas de recuperação da informação.

Os aspectos fundamentais da pesquisa sob a perspectiva da DSR, revela uma abordagem para a condução deste estudo, uma vez que proporciona uma estrutura sólida para a criação e avaliação de artefatos que visam resolver problemas práticos, com ênfase na criação e avaliação, como o *Web Crawler* desenvolvido, é um exemplo claro de como a DSR pode ser aplicada de maneira eficaz para abordar desafios complexos na CI.

Além disso, essa seção destacou a importância da compreensão dos algoritmos de motores de busca e sua aplicação centrada no usuário, o que reforça a relevância deste estudo para melhorar a experiência de busca e fornecer resultados mais relevantes. O rigor metodológico e o compromisso com a precisão e integridade dos dados ressoam consistentemente com os princípios da DSR, solidificando a credibilidade e validade deste trabalho.

A próxima seção, que trata da Metodologia, promete ser crucial para a compreensão de como os princípios da DSR foram efetivamente aplicados nesta pesquisa. Ela fornece um arcabouço detalhado sobre os métodos específicos utilizados, a abordagem adotada para a investigação e como os artefatos foram desenvolvidos e avaliados. A metodologia é, sem dúvida, um pilar central para esta pesquisa, pois irá revelar o processo pelo qual as contribuições teóricas e práticas são obtidas, demonstrando como os desafios identificados e abordados de maneira sistemática e cientificamente rigorosa. Portanto, a próxima seção lançará luz sobre a estrutura que sustenta a geração de conhecimento e a solução de problemas na área da CI.

4 METODOLOGIA

A pesquisa adotou a metodologia da *Design Science Research* (DSR), que fornece um arcabouço teórico para orientar o pesquisador na criação de conhecimento durante a concepção de artefatos. A abordagem de pesquisa integra os paradigmas da ciência do comportamento e da ciência do *design*. As técnicas de coleta de dados utilizadas incluem observação direta, análise de documentos e revisão bibliográfica, visando obter informações relevantes sobre os algoritmos dos motores de busca. O tratamento dos dados coletados é realizado por meio de análise qualitativa, que inclui categorização e codificação. No desenvolvimento e avaliação de artefatos, foi criado um exemplo de *Web Crawler* usando *Python*, explorando a interação entre automação e algoritmos de busca. Os resultados são analisados com base na abordagem da DSR, fornecendo *insights* e conclusões relevantes para a compreensão do comportamento dos algoritmos nos motores de busca. A pesquisa contribui para a comunidade científica com sua abordagem focada na construção teórica sobre o tema. Apresenta-se, a seguir, o Quadro 5. que traz a descrição sobre cada elemento da pesquisa desenvolvida neste estudo:

Quadro 5 - Elemento da Pesquisa.

Elementos	Descrição
Tipo de pesquisa	Exploratória e descritiva, com estudo de caso como estratégia.
Abordagem de pesquisa	Abordagem qualitativa
Método de pesquisa	Metodologia DSR focada em Ciência do Projeto.
Técnica de Coleta de Dados	Levantamento e Análise Documental.
Técnica de Tratamento de Dados	Procedimentos e processos utilizados para manipular, organizar, analisar e transformar dados brutos em informações úteis.
Teste como Técnica de Tratamento de Dados	Avaliação interativa dos artefatos criados para verificar eficácia e refinamento

Fonte: Elaborado por Ocante Antônio Ié (2023).

O Quadro 5 - Elemento da Pesquisa sintetiza os aspectos essenciais da pesquisa em questão. Trata-se de uma investigação de natureza exploratória e descritiva que adota uma abordagem qualitativa. A Metodologia Design Science Research (DSR) é o método de pesquisa selecionado, com foco específico em Produto final. A coleta de dados é conduzida por meio de levantamento e análise documental. No tratamento dos dados, são empregados procedimentos e processos destinados a manipular, organizar, analisar e transformar dados brutos em informações úteis, abrangendo verificações, limpezas e organização para facilitar a análise. A tabela destaca o teste como uma técnica adicional de tratamento de dados, sublinhando a

avaliação interativa dos artefatos gerados para verificar a eficácia e refinamento do processo de pesquisa.

Esses elementos colaboram de maneira integrada para estruturar o escopo e a metodologia da pesquisa. A ênfase recai sobre a exploração e descrição de dados qualitativos, respaldadas pela Metodologia DSR com foco em criação de um Produto. A coleta de dados, realizada por levantamento e análise documental, oferece uma base sólida para compreender o contexto estudado. Os procedimentos e processos aplicados no tratamento de dados desempenham um papel crucial na transformação de dados brutos em *insights* significativos, incluindo verificações de qualidade, limpezas e organização eficiente para análise. A inclusão da técnica de teste como parte do tratamento de dados destaca a importância da avaliação contínua, contribuindo para aprimorar a eficácia e refinamento dos resultados da pesquisa. Em conjunto, esses elementos proporcionam uma visão abrangente da abordagem e das estratégias empregadas na condução da pesquisa proposta.

4.1 Tipo de Pesquisa e Abordagem

O presente trabalho traz uma pesquisa de natureza exploratória e descritiva apresentando, enquanto estratégia metodológica, um estudo de caso, com abordagem qualitativa. Quanto às frentes complementares, esta pesquisa avança por meio de levantamento e análise documental, a exemplo das técnicas de mecanismos de busca na web de dados através de uso de algoritmos. A opção por essa abordagem metodológica fundamenta-se na longa tradição que recursos metodológicos de pesquisa possuem no âmbito das Ciências Sociais e Tecnológicas (Becker, 1999).

Além disso, a abordagem qualitativa dessa pesquisa permite uma análise detalhada dos dados coletados, o que é fundamental para compreender as nuances e complexidades do tema em estudo.

De acordo com Yin (2014, p. 13), a abordagem qualitativa disponibiliza uma vasta gama de métodos para a aquisição de dados e a apreensão da realidade. Como também, enfatiza a significativa relevância da abordagem qualitativa e das mencionadas técnicas na busca por uma compreensão mais profunda.

Para isso, a utilização de mecanismos de busca na *web* e algoritmos para coleta de dados é uma estratégia importante para ampliar a base de dados e garantir a precisão e

confiabilidade dos resultados. A combinação desses elementos metodológicos é fundamental para garantir a qualidade e validade da pesquisa científica e para alcançar os objetivos propostos.

4.2 Método da Pesquisa

A condução da pesquisa sobre algoritmos e sistemas de informação é crucial para assegurar a validade e confiabilidade dos resultados obtidos. Diversas metodologias de pesquisa estão disponíveis, como o método etnográfico, o histórico e o estudo de caso, conforme destacado por Egghe e Rousseau (1990). No contexto deste estudo, optou-se por seguir a metodologia de pesquisa delineada por Hevner e Chatterjee (2010) para a *Design Science Research* (DSR). Hevner e Chatterjee (2010) enfatizam que a DSR representa uma metateoria centrada na evolução da Ciência do Projeto, originada na Engenharia de Produção. Esta abordagem destaca a operacionalização por meio da aplicação de diretrizes e processos específicos para conduzir pesquisas de design científico.

Nessa abordagem, a pesquisa está alinhada ao conceito de Ciência do Projeto, também conhecida como Ciência do Artificial, conforme proposto por Simon (1996). Dentro desse contexto, a Ciência do Projeto contribui para a geração de conhecimento científico, enfocando as construções, métodos, técnicas, modelos e teorias desenvolvidos no processo de criação de artefatos.

O método adotado para estruturar esta pesquisa, fundamentado na Ciência do Projeto, segue a proposta de Hevner e Chatterjee (2010). Desta forma, o trabalho detalha os ciclos de pesquisa que conduziram à construção de protótipos dos algoritmos de motores de busca na web, visando a compreensão aprofundada dos processos desses mecanismos. A escolha deliberada foi por uma abordagem metodológica centrada na produção de artefatos, fundamentada na DSR.

Essa abordagem proporciona uma maior integração entre o meio acadêmico e as organizações, uma vez que segue uma estrutura análoga à condução de projetos de construção e aplicação em prática. Essa metodologia engloba etapas fundamentais, tais como planejamento, projeto, operação, verificação, avaliação e validação, frequentemente apoiadas por protocolos e verificações.

Ao optar pela abordagem do DSR, a pesquisa busca desenvolver soluções práticas e aplicadas para problemas específicos, por meio da criação de artefatos que são fundamentados em princípios científicos. Esse artefato tem como objetivo principal contribuir para

compreensão do comportamento dos algoritmos dos processos de automação dos motores de busca na perspectiva da CI.

4.3 Técnica de Coleta de Dados

As técnicas de coleta de dados empregadas nesta pesquisa, ancoradas na metodologia DSR e no domínio da CI, incorporam instrumentos específicos para adquirir informações pertinentes sobre os algoritmos dos motores de busca e os mecanismos de classificação.

A observação direta se destaca como uma dessas técnicas, possibilitando a obtenção de informações ao observar diretamente o comportamento dos algoritmos nos motores de busca. Para isso, foram realizadas análises detalhadas do desempenho desses algoritmos em ambientes controlados, permitindo uma compreensão aprofundada de suas características e funcionamento no contexto da pesquisa. Segundo Hevner e Chatterjee (2010), a observação direta é uma técnica valiosa na DSR, pois permite ao pesquisador coletar dados em tempo real e estudar os fenômenos de interesse de forma direta.

A análise de documentos desempenha um papel importante na coleta de dados, fornecendo embasamento teórico relevante para a pesquisa. Através da análise de documentos, como artigos científicos, relatórios técnicos e documentos normativos, é possível obter informações sobre as estruturas e o funcionamento dos algoritmos, além de compreender as teorias e os conceitos subjacentes à CI e à automação dos motores de busca.

A revisão bibliográfica é uma técnica amplamente utilizada na pesquisa científica, que consiste em examinar e analisar a literatura existente sobre um determinado assunto. Por meio desse processo, é possível identificar o estado atual do conhecimento na área da CI, compreender os avanços tecnológicos e as tendências relacionadas aos algoritmos dos motores de busca e obter *insights* teóricos relevantes para a pesquisa.

É importante ressaltar que a escolha dessas técnicas de coleta de dados está alinhada com a abordagem da DSR, que busca criar conhecimento teórico por meio do desenvolvimento e avaliação desses artefatos. Conforme Bax (2015), a DSR permite ao pesquisador criar e avaliar artefatos com base em conjecturas teóricas, fornecendo novos dados que contribuem para o avanço do conhecimento existente.

4.4 Técnica de Tratamento de Dados

No tratamento dos dados nesta pesquisa, que se debruça em metodologia DSR na área da CI, serão empregadas técnicas de análise qualitativa. Essas técnicas permitirão extrair compreensão dos dados coletados e contribuição para o aprimoramento do conhecimento teórico existente.

A análise qualitativa é uma abordagem utilizada na pesquisa em CI, especialmente na DSR, que visa compreender e interpretar fenômenos complexos. Segundo Hevner e Chatterjee (2010), a análise qualitativa permite uma compreensão profunda dos contextos e das interações entre os elementos envolvidos na pesquisa.

Na etapa de codificação, procedimento central da análise qualitativa, rótulos ou códigos são atribuídos aos dados coletados. Esse processo possibilita a identificação de temas, relações e conexões entre diferentes elementos presentes nos dados. A codificação é um procedimento interativo e reflexivo, no qual os pesquisadores interpretam e atribuem significado aos dados coletados. Os rótulos ou códigos são escolhidos cuidadosamente para refletir a essência dos dados, e a codificação é realizada de maneira sistemática para garantir a consistência e a validade dos resultados. Este processo é essencialmente iterativo, permitindo ajustes conforme a compreensão dos dados se aprofunda ao longo da análise.

Essa técnica é essencial para a compreensão dos dados e ao analisar qualitativamente os dados coletados, é possível extrair conhecimentos e informações que ajudarão a entender o comportamento dos algoritmos nos motores de busca, contribuindo para o avanço do campo da CI.

4.5 Teste como Técnica de Tratamento de Dados

No contexto do projeto, o "Teste" é utilizado como uma técnica de tratamento de dados, envolvendo a avaliação interativa dos artefatos criados durante a pesquisa. Esse processo é fundamental para verificar a eficácia e o refinamento dos resultados obtidos. Nesse caso, o teste não se refere a testes de *software*, mas sim à avaliação prática e contínua dos artefatos desenvolvidos no âmbito da DSR.

A abordagem de teste envolve a implementação e a aplicação dos algoritmos de motores de busca da *web*, conforme desenvolvidos durante a pesquisa, em situações reais ou simuladas. Essa implementação permite avaliar o desempenho dos algoritmos, a qualidade dos resultados produzidos e a sua conformidade com os objetivos definidos. Ao submeter esses algoritmos a diferentes cenários e conjuntos de dados, é possível identificar possíveis falhas, limitações ou oportunidades de aprimoramento.

A avaliação ocorre em ciclos iterativos, onde os resultados obtidos são analisados e comparados com as expectativas iniciais. Com base nessas análises, são realizados ajustes, refinamentos e otimizações nos algoritmos. Esse processo é repetido até que os artefatos alcancem os níveis desejados de desempenho, eficácia e eficiência.

Assim, o teste como técnica de tratamento de dados no projeto envolve a aplicação prática dos algoritmos desenvolvidos, seguida de análise e ajustes iterativos. Essa abordagem permite validar os resultados, garantir a qualidade dos artefatos e contribuir para o avanço do conhecimento no campo da CI, proporcionando uma compreensão maior dos comportamentos dos algoritmos de motores de busca na *web*.

A pesquisa adota uma abordagem metodológica que se baseia nos princípios da DSR, e através dessa abordagem, a pesquisa busca não apenas compreender os algoritmos de motores de busca, mas também criar soluções práticas e relevantes para a área da Ciência da Informação. Essa análise, delineada na tabela anterior sob as categorias como o comportamento de motores de busca na *web*, as formas de extração e indexação de dados, busca automatizada de informação na *web* de dados permite uma investigação profunda e sistemática dos mecanismos de busca e dos algoritmos subjacentes.

4.6 Análise SWOT

A pesquisa em CI sobre Algoritmos de Motores de Busca apresenta uma análise *Strengths, Weaknesses, Opportunities and Threats* (SWOT), essencial para compreender seu contexto e direcionar estratégias de desenvolvimento. Esta abordagem proporciona uma visão abrangente dos principais aspectos que impactam a pesquisa, identificando forças e fraquezas internas, bem como oportunidades e ameaças externas.

Tsatsou et al. (2011), realizaram uma análise SWOT para avaliar a implementação da educação a distância em universidades tradicionais. Eles usaram a tabela SWOT para identificar os pontos fortes, pontos fracos, oportunidades e ameaças dessa implementação.

Assim como para Lesk (2004) no livro "*Understanding Digital Libraries*", o autor utiliza tabelas SWOT para avaliar vários aspectos relacionados à gestão de documentos eletrônicos e bibliotecas digitais. Ele emprega essa ferramenta de análise para destacar os pontos fortes, pontos fracos, oportunidades e ameaças associadas a essas áreas.

Ambos os autores demonstram como a análise SWOT pode ser uma ferramenta valiosa para avaliar diferentes contextos, projetos e áreas de pesquisa, fornecendo uma visão abrangente que ajuda na tomada de decisões e no planejamento estratégico.

A análise SWOT foi conduzida no âmbito da pesquisa em Ciência da Informação (CI) sobre Algoritmos de Motores de Busca. Essa abordagem estratégica se revelou essencial para compreender o contexto dessa pesquisa e orientar estratégias de desenvolvimento. Ao empregar a análise SWOT, foram identificados os pontos fortes, pontos fracos, oportunidades e ameaças inerentes a esse campo específico.

Exemplificando a utilidade da análise SWOT, Tsatsou et al. (2011) aplicaram essa abordagem ao avaliar a implementação da educação a distância em universidades tradicionais, identificando os elementos críticos relacionados a essa iniciativa. Lesk (2004), em "Understanding Digital Libraries", utilizou a análise SWOT para avaliar a gestão de documentos eletrônicos e bibliotecas digitais, destacando os pontos fortes, pontos fracos, oportunidades e ameaças associadas.

Ambos os autores destacaram a versatilidade dessa ferramenta em avaliar contextos diversos, projetos e áreas de pesquisa. No contexto específico da pesquisa em questão, a análise SWOT revelou pontos fortes, como a abordagem interdisciplinar. Por outro lado, os pontos fracos estão relacionados às limitações de recursos e à complexidade inerente ao tema dos algoritmos de motores de busca. As oportunidades incluem avanços tecnológicos e seu impacto positivo na sociedade. Já as ameaças abrangem possíveis mudanças regulatórias, manutenção e atualização.

4.6.1 Pontos Fortes (*Strengths*)

Na abordagem interdisciplinar, a pesquisa em CI se beneficia ao unir conhecimentos teóricos da área com a prática da Tecnologia da Informação. Conforme destacado por Meneghini (2012), a interdisciplinaridade é fundamental para a evolução do campo, permitindo abordagens mais ricas e a solução de problemas complexos.

Em relação a metodologia, a escolha do DSR oferece um arcabouço sólido e sistemático para o desenvolvimento da pesquisa, conforme sugerido por Baranauskas et al. (2009). Isso contribui para a estruturação dos processos de pesquisa e a garantia da qualidade dos resultados.

A pesquisa incorpora técnicas, como observação direta, análise de documentos e revisão bibliográfica, demonstrando um compromisso com a obtenção de dados confiáveis e relevantes, conforme destacado por Bax (2015) ao ressaltar a importância de técnicas na pesquisa em CI. Enquanto na avaliação contínua, a técnica de tratamento de dados que envolve a avaliação interativa dos artefatos desenvolvidos garante a melhoria contínua e a validação dos

resultados, seguindo a abordagem de Borsato (2013) sobre a importância da validação constante em pesquisas científicas.

4.6.2 Pontos Fracos (*Weaknesses*)

Na limitação de recursos, as restrições de recursos financeiros e temporais podem afetar a extensão da pesquisa e a implementação de soluções práticas, conforme discutido por Packer (2010) sobre os desafios de pesquisa no Brasil devido à falta de investimento.

Com a complexidade do tema, a pesquisa aborda um tema complexo relacionado a algoritmos de motores de busca, exigindo um alto nível de especialização e compreensão, como apontado por Lara et al. (2019) ao destacar a necessidade de pesquisadores altamente capacitados na área.

Enquanto aos desafios éticos, essas questões, como privacidade e viés algorítmico, são desafios enfrentados pela pesquisa, alinhando-se com as preocupações de Barbosa (2018) sobre a ética na pesquisa em CI e Tecnologia.

4.6.3 Oportunidades (*Opportunities*)

Nos avanços tecnológicos, o campo de algoritmos de motores de busca está em constante evolução, proporcionando oportunidades para pesquisas inovadoras e contribuições significativas, conforme ressaltado por Guimarães (2020) ao destacar a rápida transformação tecnológica na área.

Abordando o impacto na sociedade, a pesquisa pode contribuir para uma compreensão mais profunda dos algoritmos de busca usados em plataformas *on-line*, ajudando a mitigar desafios como a disseminação de desinformação e a privacidade dos usuários, alinhando-se com as preocupações de Cunha (2017) sobre o papel da CI na sociedade.

4.6.4 Ameaças (*Threats*)

Considerando a viabilidade técnica dos algoritmos, a pesquisa identificou desafios que podem comprometer sua eficácia em ambientes reais. Dificuldades na manutenção e atualização dos algoritmos, confiabilidade dos dados de treinamento, interoperabilidade, possíveis ataques adversários, escalabilidade, exigências de recursos, privacidade e segurança, aceitação do usuário e conformidade regulatória foram destacadas como aspectos críticos a serem

considerados no desenvolvimento e aplicação prática dos algoritmos. Essas considerações refletem a complexidade e os diversos desafios associados à implementação bem-sucedida de algoritmos de motores de busca na área de CI.

Quadro 6 - Principais pontos na análise SWOT.

Aspectos	Pontos Fortes 	Pontos Fracos 	Oportunidades 	Ameaças 
Internos	Abordagem Interdisciplinar	Limitação de Recursos	-----	-----
	Metodologia Robusta	Complexidade do Tema	-----	-----
	Utilização de Técnicas Avançadas	Desafios Éticos	-----	-----
	Avaliação Contínua	-----	-----	-----
Externo	Avanços Tecnológicos	limitação de recursos	Impacto na Sociedade	Mudanças Regulatórias, manutenção e atualização

Fonte: Elaborado por Ocante António Ié (2023).

Essa análise SWOT oferece uma visão clara da pesquisa em CI sobre Algoritmos de Motores de Busca, destacando áreas em que a pesquisa pode ser aprimorada e onde estratégias podem ser implementadas. Uma maneira eficaz de aproveitar esses *insights* e enfrentar os desafios é desenvolver um produto que aproveite os pontos fortes e as oportunidades identificados. Nesse contexto, o próximo passo a ser abordado será a descrição desse produto, que visa contribuir significativamente para a compreensão dos algoritmos de busca na *web* e seu impacto na sociedade.

5 PRODUTO

Nesta seção, abordaremos o produto que emergiu como resultado desta pesquisa, que se concentra em solucionar o problema central identificado na pesquisa em CI sobre Algoritmos de Motores de Busca, direcionando com a necessidade premente de compreender de forma aprofundada os processos de automação desses motores e os algoritmos que os conduzem.

Para atender à necessidade de compreensão mais ampla dos mecanismos de busca na web, desenvolvemos um Web Crawler como uma peça central deste estudo. O Web Crawler, construído usando a linguagem de programação Python, é uma ferramenta projetada para automatizar a coleta de dados na web de maneira precisa e eficaz.

Para explicar em termos mais acessíveis, um Web Crawler é como um "robô" da internet que percorre páginas da web em busca de informações específicas. Imagine-o como um explorador digital que segue links entre páginas da web, extrai dados relevantes e os organiza para análise. O funcionamento básico envolve três etapas:

Identificação de Páginas para Exploração: O processo de identificação de páginas para exploração inicia-se com o Web Crawler visitando uma ou mais páginas iniciais, servindo como ponto de partida. Nesse estágio, o Crawler identifica os links contidos nessas páginas e coleta esses links para acessar outras páginas. Para ilustrar esse procedimento, utilizamos como exemplo o estudo que visa coletar informações da Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (Brapci). O link específico dessa base foi configurado no Crawler para a extração de informações durante o processo.

Exploração e Coleta de Dados: O Crawler segue os links identificados, acessa as páginas correspondentes e extrai as informações relevantes. Essas informações podem incluir texto, imagens, links, entre outros, dependendo dos objetivos da pesquisa.

Organização e Armazenamento: Após coletar os dados, o Web Crawler os organiza de maneira estruturada para facilitar a análise subsequente. Os dados podem ser armazenados em um formato adequado para análise, como um banco de dados ou arquivo.

É importante destacar que o desenvolvimento do Web Crawler requer conhecimentos técnicos em programação, neste estudo, selecionamos a linguagem Python. No entanto, sua utilização na pesquisa permite uma coleta eficiente de dados em larga escala na

web, contribuindo para uma compreensão mais profunda dos mecanismos de busca e suas nuances.

Este artefato é projetado para seguir *links* e indexar conteúdo em páginas da *web*, replicando, assim, o comportamento de motores de busca. Sua capacidade de explorar a *web* sistematicamente permite a coleta de dados abrangentes e relevantes para análise.

A aplicação prática deste *Web Crawler* é de considerável importância, uma vez que ele permite a investigação minuciosa do funcionamento dos algoritmos de busca em motores específicos. Através da implementação e execução do *Web Crawler*, conseguimos coletar dados que possibilitaram uma análise detalhada do comportamento desses algoritmos em diversos contextos e cenários de busca.

Este artefato não apenas enriquece nossa compreensão dos processos de automação dos motores de busca, como também tem o potencial de impactar positivamente várias áreas da CI. Profissionais, pesquisadores e estudiosos podem utilizar essa ferramenta para conduzir investigações em larga escala relacionadas aos mecanismos de busca na *web*, melhorando, assim, a qualidade e a precisão das análises.

O *Web Crawler* desenvolvido como resultado desta pesquisa representa uma contribuição significativa para a Ciência da Informação (CI). Essa ferramenta oferece uma solução prática e concreta para uma lacuna identificada na área, preenchendo uma necessidade crucial de conhecimento.

A lacuna identificada refere-se à limitação no acesso eficiente e preciso a dados relevantes na *web* para estudos em CI, especialmente no contexto da pesquisa sobre algoritmos de motores de busca. O problema técnico associado a essa lacuna envolve a coleta manual de dados, que pode ser demorada, sujeita a erros e limitada em escala.

O *Web Crawler* surge como uma resposta a esse desafio, automatizando o processo de coleta de dados na *web*. Ao fazê-lo, supera as limitações associadas à abordagem manual, proporcionando uma maneira eficaz de extrair informações necessárias para análises em CI. Dessa forma, o *Crawler* não apenas simplifica o processo, economizando tempo e recursos, mas também melhora a precisão e a abrangência da coleta de dados, ampliando as possibilidades de pesquisa na área.

Além disso, valida a aplicação eficaz da metodologia DSR na pesquisa em CI. Através da criação deste *Web Crawler*, demonstramos a viabilidade de utilizar abordagens de *design* científico para desenvolver soluções baseadas em princípios sólidos.

O produto resultante desta pesquisa, o *Web Crawler*, é uma solução concreta para o problema da compreensão dos algoritmos de motores de busca na *web*. Sua aplicação prática

e o impacto potencial que oferece na pesquisa em CI destacam sua relevância e contribuição para o avanço do conhecimento nesta área.

5.1 Documentação

O presente projeto de algoritmo de motor de busca, consiste na construção de uma aplicação *web* de *web scraping* utilizando o *framework Flask*. Esta aplicação visa permitir aos usuários realizar pesquisas por meio de um termo de interesse, executando posteriormente um processo de *web scraping* para capturar informações relevantes de páginas relacionadas ao termo de pesquisa. O algoritmo é composto pelas seguintes etapas:

- **Importação de Bibliotecas:** Inicialmente, são importadas as bibliotecas necessárias para o desenvolvimento da aplicação. A biblioteca *Flask* é empregada para criar a estrutura da aplicação *web*, enquanto as bibliotecas *render_template* e *request* são utilizadas, respectivamente, para renderizar os *templates* HTML e lidar com as requisições HTTP. Além disso, as funções *motor_de_busca* e *obter_urls_capturadas* do arquivo *webscrapping* são importadas para realizar as operações de busca e coleta de URLs;
- **instanciação do Flask:** Uma instância da aplicação *Flask* é criada, fornecendo a base para a construção das rotas e funcionalidades da aplicação;
- **Rota Principal ('/')**: É definida a rota principal da aplicação, que corresponde à página inicial. Esta rota renderiza o *template index.html*, proporcionando ao usuário um formulário de pesquisa;
- **Rota '/search':** A rota *'/search'* é definida para lidar com as requisições do formulário de pesquisa. Utilizando o método POST, esta rota é acionada quando o usuário submete um termo de pesquisa;
- **Função *search()*:** Ao acessar a rota *'/search'*, a função *search()* é executada. Essa função obtém o termo de pesquisa digitado pelo usuário e as URLs capturadas durante buscas anteriores;
- **Execução da função de busca (*motor_de_busca*):** A função *motor_de_busca* é invocada, recebendo como argumentos o termo de pesquisa e as URLs capturadas. O objetivo é executar uma busca nas páginas relacionadas ao termo e coletar informações relevantes;
- **Formatação dos Resultados:** Após a execução da busca, os resultados são armazenados em uma lista chamada *resultados_formatados*. Um cabeçalho indicando o termo de pesquisa é adicionado como primeiro elemento dessa lista;

- Percorrendo os Resultados: A lista de resultados é percorrida, verificando se cada elemento é um dicionário contendo informações detalhadas ou apenas uma *string* representando um cabeçalho. Caso seja um dicionário, ele é adicionado à lista `resultados_formatados`;
- Apresentação dos Resultados: Os resultados formatados são impressos no console, distinguindo entre os cabeçalhos e as informações detalhadas. Além disso, o *template results.html* é renderizado, passando a lista `resultados_formatados` como parâmetro para exibição na página;
- Execução da Aplicação *Flask*: Quando o arquivo é executado diretamente, a aplicação *Flask* é iniciada, permitindo que os usuários acessem a *interface web*, realizem pesquisas e visualizem os resultados.

Quadro 7 - Sínteses das etapas da construção.

A elaboração do algoritmo de motor de busca, como descrito nas etapas detalhadas, é fundamental para atingir os objetivos da pesquisa. Cada fase desempenha um papel crucial na construção da aplicação web de web scraping. A importação de bibliotecas, a instanciação do *Flask* e a definição das rotas constituem a estrutura básica da aplicação. A rota principal e a rota de pesquisa fornecem interfaces intuitivas para os usuários, enquanto a função de busca (`motor_de_busca`) realiza operações essenciais, explorando páginas relacionadas ao termo de pesquisa e coletando informações.

A formatação dos resultados e a apresentação na interface web, juntamente com a execução da aplicação *Flask*, tornam o algoritmo funcional e acessível. A síntese das etapas destaca a importância da busca, coleta, formatação e apresentação de dados para proporcionar aos usuários uma experiência de pesquisa eficiente e informativa. Em resumo, o processo de construção delineado é fundamental para o desenvolvimento de um algoritmo de busca, alinhado com os objetivos da pesquisa em Ciência da Informação.

PROJETO DE ALGORITMO DE MOTOR DE BUSCA

SÍNTESES DAS ETAPAS DA CONSTRUÇÃO

IMPORTAÇÃO DE BIBLIOTECAS

- Flask: criação da estrutura da aplicação web.
- `render_template` e `request`: renderização de templates HTML e tratamento de requisições HTTP.
- Funções `motor_de_busca` e `obter_urls_capturadas` do arquivo `webscrapping`: realização de operações de busca e coleta de URLs.

INSTANCIACIÃO DO FLASK

Criação de uma instância do Flask, fornecendo a base para a construção das rotas e funcionalidades da aplicação.

ROTA PRINCIPAL ('/')

Definição da rota principal da aplicação, correspondendo à página inicial. Essa rota renderiza o template `index.html`, apresentando ao usuário um formulário de pesquisa.

ROTA '/SEARCH':

Definição da rota `/search` para lidar com as requisições do formulário de pesquisa, acionada quando o usuário submete um termo de pesquisa.

FUNÇÃO SEARCH()

Execução da função `search()` ao acessar a rota `/search`, obtendo o termo de pesquisa digitado pelo usuário e as URLs capturadas durante buscas anteriores.

EXECUÇÃO DA FUNÇÃO DE BUSCA (MOTOR_DE_BUSCA)

Invocação da função `motor_de_busca`, recebendo o termo de pesquisa e as URLs capturadas, realizando a busca nas páginas relacionadas ao termo e coletando informações relevantes.

FORMATACÃO DOS RESULTADOS

Armazenamento dos resultados em uma lista chamada `resultados_formatados`, com um cabeçalho indicando o termo de pesquisa como primeiro elemento.

PERCORRENDO E APRESENTANDO OS RESULTADOS

Percorrendo a lista de resultados, adicionando à lista `resultados_formatados` os dicionários com informações detalhadas. Os resultados são apresentados no console e no template `results.html`.

EXECUÇÃO DA APLICAÇÃO FLASK

Quando o arquivo é executado diretamente, a aplicação Flask é iniciada, permitindo aos usuários acessar a interface web, realizar pesquisas e visualizar os resultados.

Fonte: Elaborado por Ocante António Ié (2023).

A demonstração visual apresentada na imagem acima ilustra de forma clara o fluxo e as etapas da construção do algoritmo de motor de busca proposto. A tabela detalhada, juntamente com o esquema visual, proporciona uma visão abrangente do processo desde a importação das bibliotecas até a execução da aplicação *Flask*. Essa representação visual facilita a compreensão do funcionamento, destacando cada componente e a interação entre eles. A clareza proporcionada pela ilustração reforça a transparência no desenvolvimento do algoritmo, tornando mais acessíveis tanto o entendimento do processo quanto a possível replicação ou extensão do projeto.

As funcionalidades da aplicação *web*:

- Pesquisa por Termo: Os usuários podem inserir um termo de interesse na página inicial, o que desencadeia a busca por informações relacionadas;
- *Web Scraping*: O processo de *web scraping* é realizado nas páginas relevantes ao termo de pesquisa, coletando informações que são posteriormente apresentadas aos usuários;
- Exibição de Resultados: Os resultados da *web scraping* são apresentados de forma formatada na página *results.html*, permitindo que os usuários visualizem as informações de maneira organizada;
- Pré-requisitos: Antes de executar a aplicação, é necessário atender aos seguintes pré-requisitos a respeito da Versão 3.7 ou superior do *Python* instalada;
- Bibliotecas *Python*: *Flask*, *requests*, *beautifulsoup4*, *mysql-connector-python*, *babel*, *nlTK*.

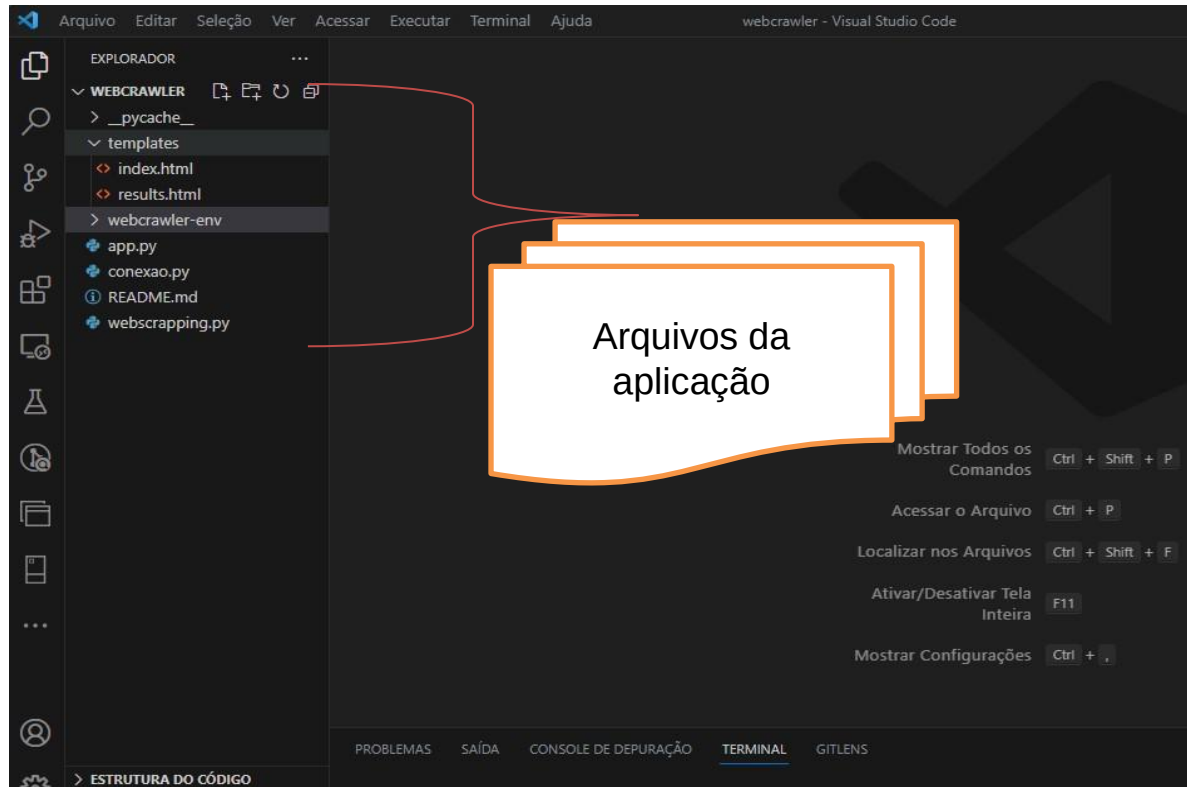
5.2 Descrição da Estrutura dos Arquivos

A estrutura de arquivos do projeto é organizada da seguinte forma (Figura 3):

- *app.py*: Arquivo principal da aplicação, contendo as rotas e a lógica de execução;
- *Templates/*: Diretório contendo os *templates* HTML;
- *index.html*: *Template* para a página inicial com o formulário de pesquisa;
- *results.html*: *Template* para a página de resultados com os dados da *web scraping*;

- *webscraping.py*: Arquivo contendo a função de *web scraping* e a obtenção das URLs capturadas.

Figura 3 – Estrutura de Arquivo.



Fonte: Elaborado por Ocante António Ié (2023).

5.3 Execução da Aplicação

Para executar a aplicação, siga os passos abaixo:

- Baixe e instale o Python a partir do site oficial: <https://www.python.org/downloads/>
- Crie uma virtual *environment* (*venv*) usando o comando: `python -m venv webcrawler-env`;

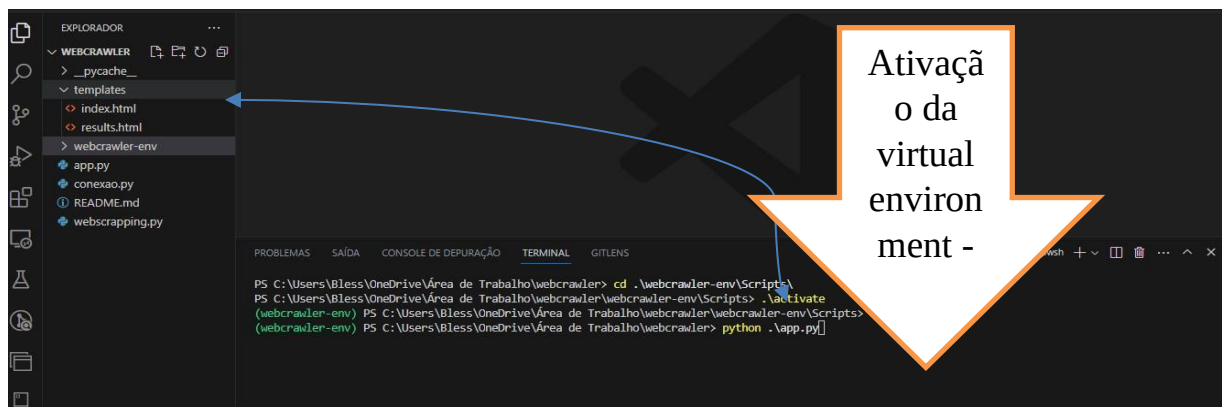
Ative a virtual environment:

- No *Windows*: `.\webcrawler-env\Scripts\activate`;
- No *macOS/Linux*: `source webcrawler-env/bin/activate`;

Instale as bibliotecas necessárias com os comandos (Figura 4):

- `pip install beautifulsoup4`
- `pip install requests`
- `pip install mysql-connector-python`
- `pip install babel`
- `pip install nltk`
- `pip install flask`

Figura 4 -Ativação de *Environment*.



Fonte: Elaborado por Ocante António Ié (2023).

Virtual *Environment* (*venv*) - é um aplicativo em rede que permite que um usuário interaja com o ambiente de computação. Permite o compartilhamento de documentos baseados na *web*.

5.4 Banco de Dados

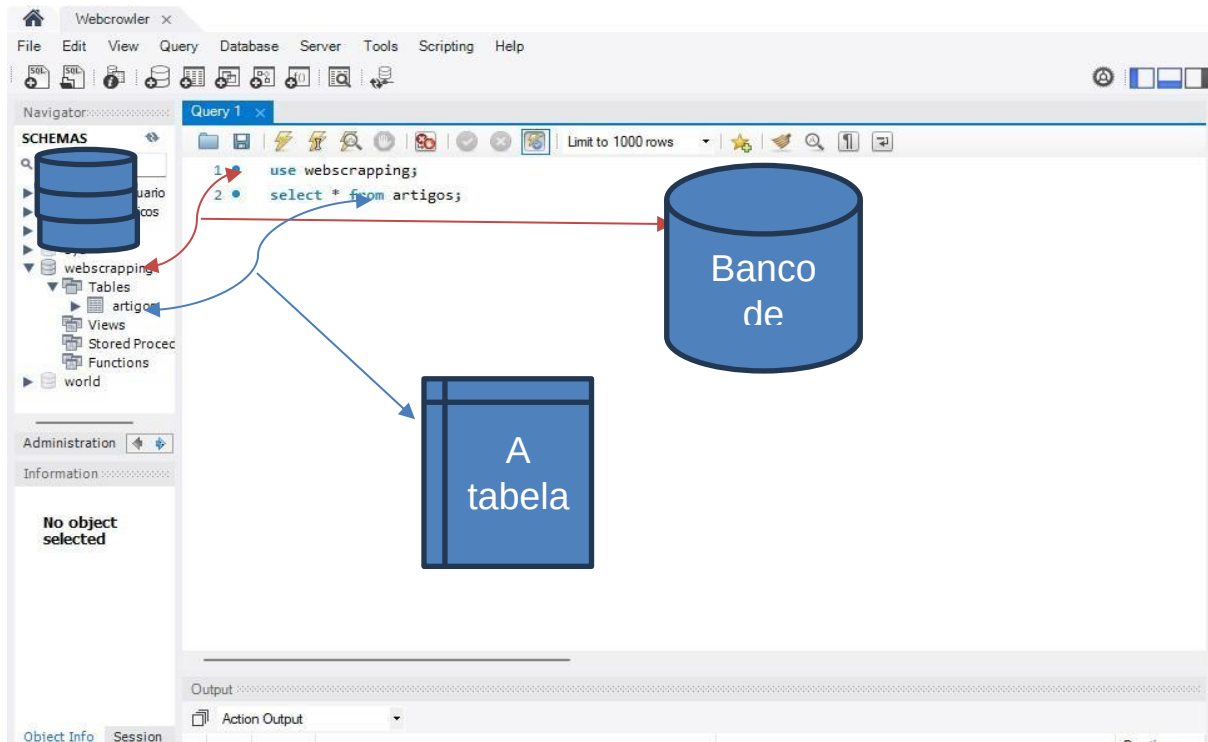
Bancos de dados ou bases de dados são conjuntos de arquivos relacionados entre si, podendo conter registros sobre pessoas, lugares ou informações em geral. São coleções organizadas de dados que se relacionam ou não, de forma a armazenar informações.

No projeto, é usado o sistema MySQL³, que serve para gerenciamento de banco de dados, utiliza a linguagem SQL como interface. E é usada a *MySQL Workbench*, que por sua vez, é uma ferramenta de *design* de banco de dados visual que integra desenvolvimento SQL,

³ Disponível em: <https://www.mysql.com/>. Acesso em: 17 jun. 2023.

administração, design de banco de dados, criação e manutenção em um único ambiente de desenvolvimento integrado para o sistema de banco de dados MySQL (Figura 5).

Figura 5 – Banco de Dados



Fonte: Elaborado por Ocante António Ié (2023).

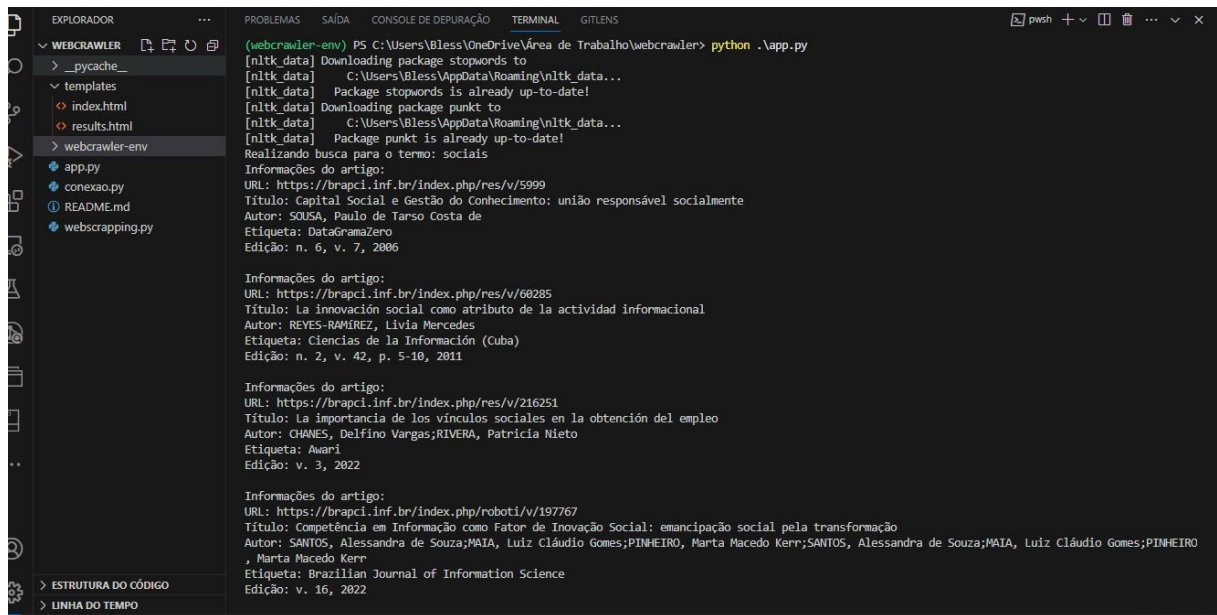
Esse banco, contém uma tabela, chamado “artigos”. Nestas tabelas, os dados são organizados de maneira lógica em um formato de linha-e-coluna semelhante ao de uma planilha. Cada linha representa um registro exclusivo e cada coluna representa um campo no registro.

5.5 Execução do Arquivo *App.Py* para iniciar a Aplicação *Flask*

Ao seguir essas etapas, a aplicação estará disponível para acesso por meio de um navegador *web*, permitindo que os usuários realizem pesquisas e visualizem os resultados obtidos por meio da *web scraping*.

O algoritmo, processa, faz a busca, coleta e indexa as informações pelo terminal (Figura 6).

Figura 6 – Extração de Informação Via Console.



```
(webcrawler-env) PS C:\Users\Bless\OneDrive\Área de Trabalho\webcrawler> python .vapp.py
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\Bless\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\Bless\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
Realizando busca para o termo: sociais
Informações do artigo:
URL: https://brapci.inf.br/index.php/res/v/5999
Título: Capital Social e Gestão do Conhecimento: união responsável socialmente
Autor: SOUSA, Paulo de Tarso Costa de
Etiqueta: DataGramazero
Edição: n. 6, v. 7, 2006

Informações do artigo:
URL: https://brapci.inf.br/index.php/res/v/60285
Título: La innovación social como atributo de la actividad informacional
Autor: REYES-RAWIREZ, Livia Mercedes
Etiqueta: Ciencias de la Información (Cuba)
Edição: n. 2, v. 42, p. 5-10, 2011

Informações do artigo:
URL: https://brapci.inf.br/index.php/res/v/216251
Título: La importancia de los vínculos sociales en la obtención del empleo
Autor: CHANES, Delfino Vargas;RIVERA, Patricia Nieto
Etiqueta: Awari
Edição: v. 3, 2022

Informações do artigo:
URL: https://brapci.inf.br/roboti/v/197767
Título: Competência em Informação como Fator de Inovação Social: emancipação social pela transformação
Autor: SANTOS, Alessandra de Souza;MAIA, Luiz Cláudio Gomes;PINHEIRO, Marta Macedo Kerr;SANTOS, Alessandra de Souza;MAIA, Luiz Cláudio Gomes;PINHEIRO, Marta Macedo Kerr
Etiqueta: Brazilian Journal of Information Science
Edição: v. 16, 2022
```

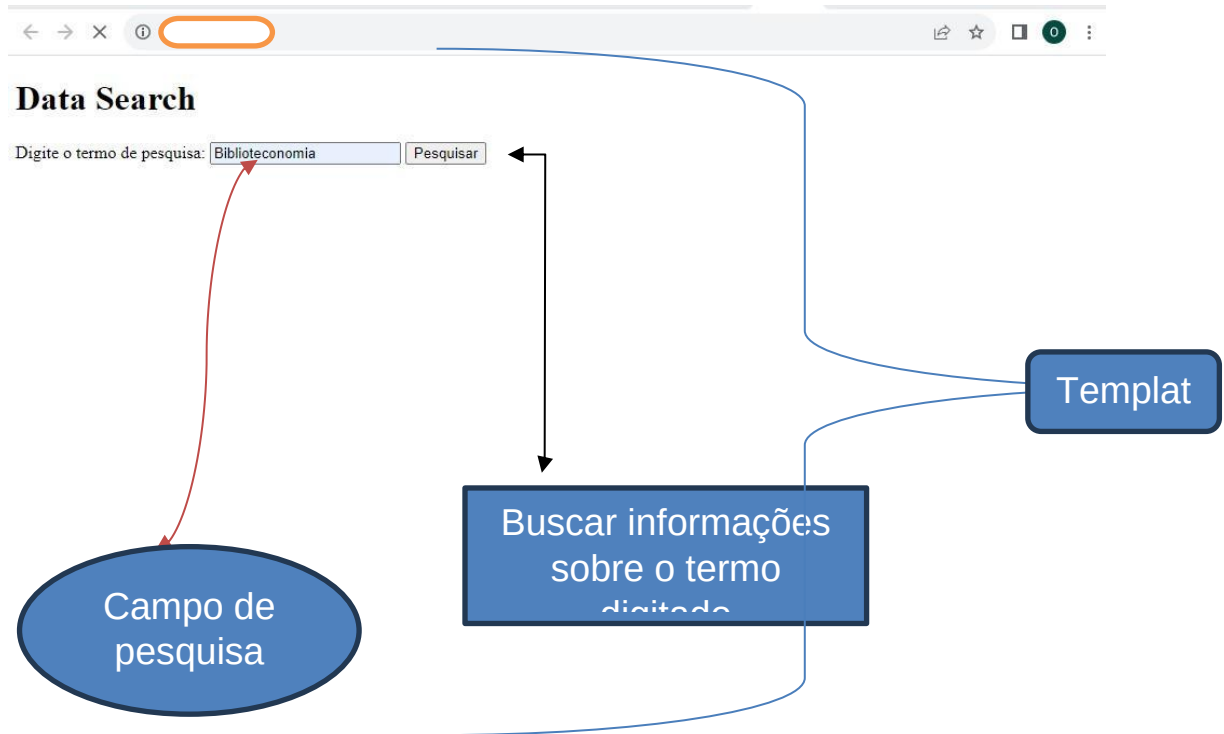
Fonte: Elaborado por Ocante Antônio Ié (2023).

A análise do algoritmo descrito no contexto da CI e Gestão de Informação destaca a aplicação prática de tecnologias e técnicas para coletar, processar e apresentar informações de maneira útil e organizada.

5.6 Etapas do Processo de Teste para o Projeto

Nesta etapa é desenvolvido um plano detalhado que descreve as estratégias, cenários e casos de teste que serão usados para avaliar o projeto (Figura 7).

Figura 7 - Template da aplicação – Página de pesquisa (Data Search).



Fonte: Elaborado por Ocante António Ié (2023).

Para que esta página seja exibida, ao iniciar pelo navegador através de <http://localhost:5000>, ele indica que um servidor *web* está sendo executado localmente (no seu próprio computador) e está ouvindo conexões na porta 5000 do endereço IP, que é um endereço reservado para a máquina local, também conhecido como "*localhost*". Isso geralmente acontece quando você está executando um aplicativo ou serviço *web* em seu computador.

5.7 Verificar se a aplicação *web* de *web scraping* retorna resultados precisos e formatados corretamente

Para o cenário, é necessário realizar uma busca com o termo "Biblioteconomia" e verificar se os resultados obtidos contêm informações relevantes.

Nos casos de teste pesquisar o termo, verificar se os resultados exibem um cabeçalho com o termo de pesquisa e analisar se os detalhes dos resultados contêm informações verdadeiras.

A execução dos testes, para testes unitários, testes da função `motor_de_busca`, verificar se a função de busca retorna resultados esperados para um termo de pesquisa específico.

Verificar se o resultado está formatado e organizado em: Autor; Etiqueta; Edição; Link; Disponibilizado; Resumo (Figuras 8 a 11).

Figura 8 – Resultados das buscas via navegador – 1.

Resultados da Pesquisa

Termo 'Biblioteconomia' encontrado em 10 páginas:

- **O código de ética dos bibliotecários brasileiros e as lições socráticas**

Autor: FARIAS, Vera Cristina de Sousa;FREIRE, Isa Maria

Etiqueta: Revista ACB: Biblioteconomia em Santa Catarina

Edição: n. 1, v. 27, p. 1-20, 2022

Link: <https://brapci.inf.br/index.php/res/v/197774>

Disponibilizado: 14/06/2022

Resumo: Aborda o bibliotecário enquanto sujeito que precisa de princípios éticos na sua prática, para crescer como ser humano e profissional, melhorar a qualidade do seu trabalho e aprimorar o relacionamento interpessoal. Considera as relações entre o Novo Código de Ética e Deontologia do Bibliotecário (2018), que apresenta normas que orientam a conduta profissional, e as Lições Socráticas, enquanto princípios importantes para a vida pessoal e instrumentos de promoção de bem-estar social. Apresenta como fundamentação teórica as ideias de Platão (2003) sobre o conhecimento; a visão de Targino (2010) sobre Ranganathan; as abordagens de Stumpf (2019) e Rasche (2005) sobre o Código de Ética dos Bibliotecários; as virtudes de Comte-Sponville (1999); e as virtudes dos bibliotecários segundo Freire, Silva e Nascimento (2015). Compartilha pesquisa bibliográfica realizada a partir de descritores temáticos de busca e abordagem qualitativa, preocupando-se com o sentido e intencionalidade dos textos. Como resultado, conclui que há correlação entre as abordagens dos autores mencionados, o novo Código de Ética e Deontologia da Biblioteconomia (2018) e as Lições Socráticas de Ismael (2004). (pt-BR)3

Fonte: Elaborado por Ocante António Ié (2023).

Figura 9 - Resultados das buscas via navegador -2.

- **Ética profissional em Ciência da Informação: uma análise de códigos de ética de arquivistas e bibliotecários**

Autor: EVANGELISTA, Isadora Victorino; SILVA, Andrieli Pachu da; GUIMARÃES, José Augusto Chaves

Etiqueta: Páginas A&B, Arquivos e Bibliotecas (Portugal)

Edição: n. 13, p. 3-12, 2020

Link: <https://brapci.inf.br/index.php/res/v/145240>

Disponibilizado: 03/08/2020

Resumo: A ética, quando relacionada à atuação profissional, visa a refletir sobre os valores que guiam o comportamento de determinada categoria, encontrando materialidade em documentos como os códigos de ética ou de deontologia, aspecto que atinge os profissionais da informação, nomeadamente os arquivistas e bibliotecários, por serem profissões que possuem exercício regulamentado em muitos países, como o Brasil. Nesse sentido, buscou-se analisar e comparar os documentos em âmbito nacional e internacional desses profissionais, em aspectos relativos a seu órgão produtor, ano de criação, denominação formal e tipos de sanções previstas. Foram recuperados 50 documentos, em inglês, italiano, francês, português e espanhol, sendo 17 para arquivistas, 30 para bibliotecários e 3 para ambos. Os códigos, em sua maioria, são provenientes de conselhos ou associações profissionais, criados no decorrer das últimas quatro décadas com denominações que variam entre código de ética, código de deontologia, código de conduta, código de ética e conduta, código de prática profissional e declaração sobre ética profissional. As sanções foram encontradas em 13 documentos, o que descaracteriza os outros 37 códigos enquanto um documento regulamentador. Os resultados levam a concluir pela necessidade de estudos mais aprofundados sobre ética e deontologia nesses dois campos profissionais no momento de se elaborarem códigos dessa natureza, de modo a garantir a necessária coesão e especificidade para o exercício profissional. Palavras-chave: Arquivologia; Biblioteconomia; Códigos de Ética; Deontologia; Ética Profissional. (pt-PT)3

Fonte: Elaborado por Ocante António Ié (2023).

Figura 10 - Resultados das buscas via navegador -3.

- **O contributo do profissional da informação na curadoria digital: contribution of the information professional in digital curation**

Autor: OLIVEIRA, Alcinete Maquine Carvalho; ROSA, Ana Isabel Sousa; GOMES, Rosângela da Silva

Etiqueta: Prisma.com (Portugal)

Edição: v. 45, 2021

Link: <https://brapci.inf.br/index.php/res/v/168371>

Disponibilizado: 29/11/2021

Resumo: O profissional da informação ganha cada vez mais espaço de atuação na atual sociedade da informação. As tecnologias de informação e comunicação aparecem para facilitar a realização dos trabalhos em meio digital. Com o crescente aumento da informação, seja digitalizada ou nascida em meio digital, tornou-se urgente o cuidado com os objetos digitais. A curadoria digital aparece como uma alternativa no cuidado dos objetivos digitais, garantindo que estes tenham um tempo maior de vida e possa ser utilizado. Acredita-se que as atividades relacionadas a curadoria digital tem exigido constantes atualizações e outras competências para o profissional da informação. Logo pretende-se saber: quais as práticas do profissional da informação que contribuem para a curadoria digital? Objetivo Geral: identificar quais tipologias do profissional da informação que realiza atividades pertinentes a curadoria digital. Objetivos específicos: destacar as literaturas que tratam sobre o profissional da informação na curadoria digital; verificar as competências necessárias para o profissional da informação atuar como curador digital; especificar os aspectos semelhantes e diferentes nas tipologias do profissional da informação. Resultados: A maioria dos curadores digitais são profissionais bibliotecários, arquivistas ou museólogos, com formação em mestrado nas áreas de biblioteconomia, gestão de registros, arquivistas ou em ciência da informação. Poucos são especializados ou tem mestrado em curadoria digital ou curadoria de dados. Algumas das atividades que o curador digital realiza inclui os seguintes serviços: criação de metadados; acesso, coleta e manipulação de conteúdo; descartes; manutenção de recursos; desenvolver políticas e gerenciamento de planos de dados e procedimentos para conteúdos digitais. Conclusão: A curadoria digital é importante para garantir a vida longa e integridade do objeto digital, para isso as organizações adaptam a sua realidade, modelos de ciclo de vida para atuar junto a estes objetos digitais. Percebe-se que a formação especializada para curador digital, está diretamente ligada as áreas de Biblioteconomia e Ciência da Informação. Destaca-se nesta pesquisa a tipologia do profissional da informação: o bibliotecário, que mais realiza atividades relacionadas a curadoria digital. (pt-PT)³

Fonte: Elaborado por Ocante António Ié (2023).

Figura 11 - Título Resultados das buscas via navegador - 4.

- **O mapeamento da produção editorial italiana traduzida no brasil por meio da descrição bibliográfica**

Autor: SILVA, José Fernando Modesto da; WATAGHIN, Lucia; DEGANI, Francisco José Saraiva

Etiqueta: Informação & Tecnologia

Edição: n. 1, v. 3, p. 66-82, 2016

Link: <https://brapci.inf.br/index.php/res/v/41109>

Disponibilizado: 05/06/2018

Resumo: Atualmente, devido ao grande desenvolvimento das comunicações, a presença da cultura e das artes italianas no Brasil está em fase de crescente intensificação e a quantidade de obras italianas que circulam no Brasil é cada vez maior. Assim, a proposta desta comunicação é apresentar a descrição bibliográfica aplicada na construção de uma base de dados bibliográficos, resultante de mapeamento editorial da literatura italiana traduzida no Brasil a partir de 1951. Dá-se destaque aos elementos descritivos utilizados, ajustados para esta finalidade, com base também na necessidade dos pesquisadores envolvidos. Na descrição do projeto verifica-se a possibilidade de colaboração efetiva gerada entre as áreas de Biblioteconomia e de Letras. Também evidencia-se a contribuição que um método tradicional como a catalogação pode ser aplicada para subsidiar a construção de uma base bibliográfica da produção editorial para fins científicos. (pt-BR)³

Fonte: Elaborado por Ocante António Ié (2023).

5.8 Teste de Integração Entre Componentes

Verificar se a aplicação *Flask* interage adequadamente com as funções de *web scraping* e o banco de dados.

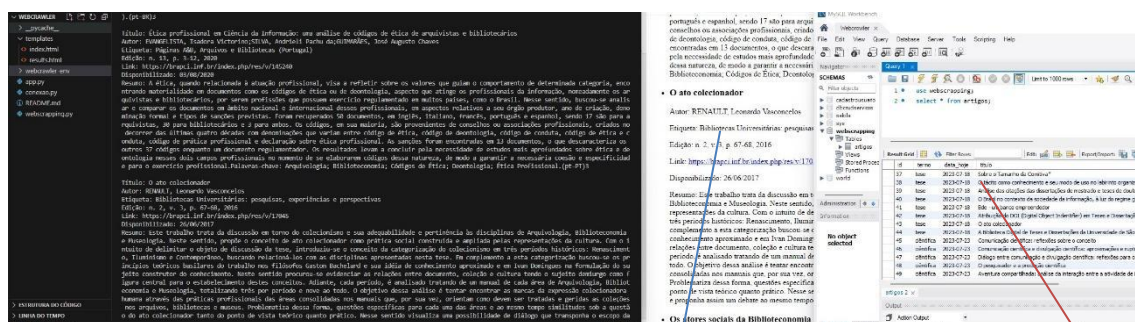
Para teste de armazenamento no banco de dados, verificar se as informações coletadas estão sendo armazenadas corretamente no banco de dados.

Dessa forma, para resultado esperado, confirmar se a busca é bem-sucedida e os resultados são exibidos. Verificando se os resultados da *web scraping* são precisos e relevantes.

Sobre a comparação de resultados esperados e reais, comparar os resultados obtidos durante os testes com as expectativas iniciais e avaliar o desempenho da aplicação sob carga de usuários.

Para a validação de dados, validação de dados inseridos, testar a aplicação inserindo dados incorretos ou inválidos nos formulários (Figura 12 e 13).

Figura 12 – Resultados Filtrados e Armazenadas.



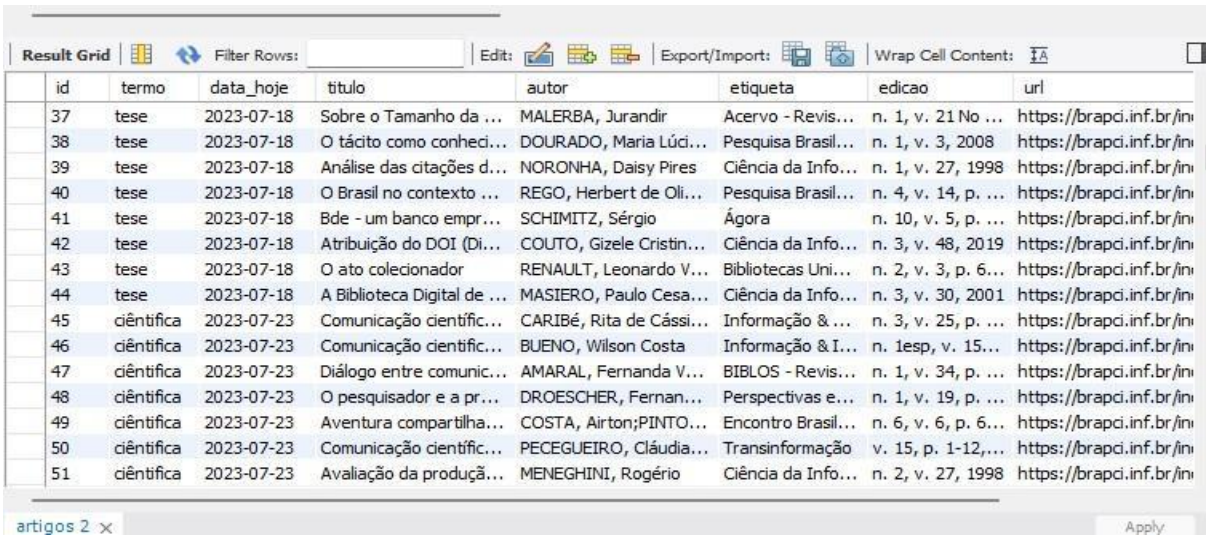
Informações no terminal mostrando o comportamento dos algoritmos de busca

As informações exibidas no terminal, são os mesmos resultados apresentados na aplicação web.

As informações coletadas são armazenadas no banco automaticamente.

Flask interage adequadamente com as funções de *web scraping* e o banco de dados. Os dados sendo coletados, indexados e armazenados no banco simultaneamente.

Fonte: Elaborado por Ocante António Ié (2023).

Figura 13 – Classificação e Organização de Informações no Banco de Dados.


id	termo	data_hoje	titulo	autor	etiqueta	edicao	url
37	tese	2023-07-18	Sobre o Tamanho da ...	MALERBA, Jurandir	Acervo - Revis...	n. 1, v. 21 No ...	https://brapci.inf.br/in
38	tese	2023-07-18	O tácito como conheci...	DOURADO, Maria Lúci...	Pesquisa Brasil...	n. 1, v. 3, 2008	https://brapci.inf.br/in
39	tese	2023-07-18	Análise das citações d...	NORONHA, Daisy Pires	Ciência da Info...	n. 1, v. 27, 1998	https://brapci.inf.br/in
40	tese	2023-07-18	O Brasil no contexto ...	REGO, Herbert de Oli...	Pesquisa Brasil...	n. 4, v. 14, p. ...	https://brapci.inf.br/in
41	tese	2023-07-18	Bde - um banco empr...	SCHIMITZ, Sérgio	Ágora	n. 10, v. 5, p. ...	https://brapci.inf.br/in
42	tese	2023-07-18	Atribuição do DOI (Di...	COUTO, Gizele Cristin...	Ciência da Info...	n. 3, v. 48, 2019	https://brapci.inf.br/in
43	tese	2023-07-18	O ato colecionador	RENAULT, Leonardo V...	Bibliotecas Uni...	n. 2, v. 3, p. 6...	https://brapci.inf.br/in
44	tese	2023-07-18	A Biblioteca Digital de ...	MASIERO, Paulo Cesa...	Ciência da Info...	n. 3, v. 30, 2001	https://brapci.inf.br/in
45	científica	2023-07-23	Comunicação científic...	CARIBÉ, Rita de Cássi...	Informação & ...	n. 3, v. 25, p. ...	https://brapci.inf.br/in
46	científica	2023-07-23	Comunicação científic...	BUENO, Wilson Costa	Informação & I...	n. 1esp, v. 15...	https://brapci.inf.br/in
47	científica	2023-07-23	Diálogo entre comunic...	AMARAL, Fernanda V...	BIBLOS - Revis...	n. 1, v. 34, p. ...	https://brapci.inf.br/in
48	científica	2023-07-23	O pesquisador e a pr...	DROESCHER, Fernan...	Perspectivas e...	n. 1, v. 19, p. ...	https://brapci.inf.br/in
49	científica	2023-07-23	Aventura compartilha...	COSTA, Ailton;PINTO...	Encontro Brasil...	n. 6, v. 6, p. 6...	https://brapci.inf.br/in
50	científica	2023-07-23	Comunicação científic...	PECEGUEIRO, Cláudia...	Transinformação	v. 15, p. 1-12,...	https://brapci.inf.br/in
51	científica	2023-07-23	Avaliação da produçã...	MENEGHINI, Rogério	Ciência da Info...	n. 2, v. 27, 1998	https://brapci.inf.br/in

Fonte: Elaborado por Ocante António Ié (2023).

No âmbito do tema "Aparato dos Algoritmos na CI: Compreendendo os Processos de Automação dos Motores de Busca", a implementação do ciclo de Lançamento e Monitoramento de uma aplicação ilustra vividamente o potencial dessa tecnologia. Ao explorar o algoritmo de um motor de busca para efetuar a coleta, indexação, armazenamento e disponibilização de informações, essa aplicação exemplifica a integração harmoniosa entre CI e Tecnologia. O ciclo de lançamento introduz a aplicação aos usuários, enquanto o monitoramento contínuo, guiado por algoritmos de automação, aprimora continuamente a experiência do usuário ao identificar problemas emergentes e realizar ajustes dinâmicos. Esse processo ilustra a capacidade dos algoritmos na CI de criar, otimizar e manter soluções que evoluem de acordo com as necessidades em constante mudança, consolidando a importância dessa disciplina na era da automação e do avanço tecnológico.

5.9 Apresentação e Análise de Dados

Nesta pesquisa, o fenômeno comportamental investigado é a comportamento dos algoritmos nos processos de automação dos motores de busca, e o artefato desenvolvido são os algoritmos de motores de busca para facilitar a compreensão do seu funcionamento na busca, indexação e armazenamento de informações. Para embasar esse estudo, Hevner *et al.* (2004) propõem um *framework* conceitual, onde os produtos da pesquisa consistem em teorias e artefatos que devem ser justificados e avaliados, a fim de gerar conhecimento que apoie a utilização dessas teorias e artefatos.

Adicionalmente, o autor disponibiliza um conjunto de diretrizes que devem ser seguidas pelas pesquisas em sistemas de informação. A perspectiva adotada também se mostrou relevante ao estabelecer as bases para futuras explorações. À medida que esta pesquisa avança para o horizonte dos trabalhos futuros, alguns caminhos emergem como naturais desdobramentos do conhecimento adquirido. A ampliação da análise empírica, por exemplo, poderia abranger estudos de caso que examinam de perto o comportamento dos algoritmos de busca em cenários específicos, proporcionando compreensão mais tangíveis sobre seu funcionamento e possíveis melhorias.

Além disso, a questão da transparência algorítmica e da mitigação de viés nos algoritmos de busca permanece como um terreno fértil para investigações mais aprofundadas. A crescente conscientização sobre os potenciais impactos negativos desses algoritmos ressalta a necessidade de desenvolver abordagens que garantam uma classificação justa e imparcial dos resultados de busca.

O campo da educação também pode se beneficiar das descobertas desta pesquisa. O desenvolvimento de ferramentas educacionais que elucidam o funcionamento dos motores de busca para o público em geral pode contribuir para uma compreensão mais informada dos usuários, capacitando-os a tomar decisões conscientes ao interagir online.

O cruzamento desses autores revela que a pesquisa sobre algoritmos de busca, quando orientada por rigor metodológico, compreensão das necessidades dos usuários e enfoque na gestão informacional, pode gerar uma série de benefícios para a CI, empresas, governos e instituições sem fins lucrativos.

No contexto das empresas, os algoritmos de busca aprimorados podem melhorar a eficiência na RI internas e externas, otimizando a tomada de decisões e impulsionando a inovação. A gestão de informações, com base em algoritmos mais precisos, pode levar a estratégias mais bem fundamentadas e eficazes.

Para governos e instituições sem fins lucrativos, os algoritmos de busca podem desempenhar um papel crucial na disseminação de informações relevantes para o público, permitindo um maior envolvimento e transparência. Eles podem ser usados para facilitar o acesso a dados públicos, relatórios governamentais e recursos educacionais, promovendo uma sociedade mais informada e participativa.

Além disso, a pesquisa sobre algoritmos de busca pode abordar preocupações contemporâneas, como a transparência algorítmica e a minimização de viés. Ao desenvolver algoritmos mais transparentes e justos, essa pesquisa pode contribuir para um uso mais ético e

inclusivo da tecnologia, evitando discriminações injustas e favorecimentos indesejados nos resultados de busca.

A pesquisa transcende a mera exploração técnica, revelando-se como uma disciplina que, quando ancorada em princípios sólidos da CI e orientada para a prática, pode moldar positivamente a forma como interagimos com a informação na era digital. Seus impactos se estendem desde a gestão informacional até a tomada de decisões em diversos setores, contribuindo para uma sociedade mais informada, engajada e ética.

Portanto, esta pesquisa ressalta a relevância do estudo sobre algoritmos de motores de busca na CI. Guiada pelas diretrizes de DSR, este trabalho desdobrou-se em uma exploração profunda e rigorosa, que se traduz em contribuições significativas para a compreensão da recuperação da informação e sua relação com a sociedade contemporânea. O tema central, "A Interseção dos Algoritmos na CI: Desvendando os Processos de Automatização dos Motores de Busca," destaca-se como um marco, delineando um percurso que harmoniza alicerces teóricos sólidos com aplicações práticas de grande impacto. Ao projetar um olhar para o futuro, esta pesquisa acende um convite para investigações mais aprofundadas, inovações e reflexões que ampliem ainda mais o conhecimento nesse campo em constante evolução.

A análise dos dados nesta pesquisa destaca a urgência contínua de aprofundamento na compreensão dos algoritmos de busca, especialmente no que diz respeito ao comportamento desses algoritmos em motores de busca automatizados. Diante da rápida evolução tecnológica e da crescente dependência da sociedade em sistemas de informação, a investigação sobre a transparência algorítmica e a mitigação de viés nos resultados de busca assume um papel crucial. Explorações futuras podem focalizar estudos de caso mais detalhados, proporcionando uma visão mais minuciosa do funcionamento dos algoritmos em contextos específicos e, assim, enriquecendo a compreensão holística desses sistemas.

Outro aspecto relevante a ser destacado na análise dos dados é a perspectiva de desenvolver ferramentas educacionais. A capacitação dos usuários por meio de conhecimentos mais profundos sobre os algoritmos de busca pode não apenas resultar em interações mais informadas e conscientes online, mas também influenciar positivamente a tomada de decisões e o discernimento de informações. Essa abordagem não só beneficia o público em geral, mas também apresenta implicações positivas para a segurança digital e a literacia informacional em um ambiente cada vez mais digitalizado.

A análise também sublinha o papel fundamental dos algoritmos de busca em diversos setores. Nas empresas, melhorias nesses algoritmos têm o potencial de transformar a

recuperação interna e externa de informações, impactando positivamente a eficiência, inovação e tomada de decisões. No âmbito governamental e de organizações sem fins lucrativos, a implementação eficaz desses algoritmos pode ser uma ferramenta poderosa na disseminação de informações relevantes, contribuindo para uma sociedade mais transparente e participativa.

Esta análise transcende a abordagem técnica, consolidando-se como uma disciplina capaz de influenciar positivamente a interação com a informação na era digital. Sua contribuição estende-se da gestão informacional até a tomada de decisões, exercendo impacto em diversos setores da sociedade. Ao considerar as implicações éticas e sociais dos algoritmos de busca, a análise ressalta a necessidade de uma abordagem crítica e responsável no desenvolvimento dessas tecnologias.

Portanto, a análise destaca que este estudo sobre os algoritmos de motores de busca na CI não apenas abre portas para investigações mais profundas, mas também instiga inovações e reflexões contínuas. Ao navegar por esse campo em constante evolução, a análise aponta para a importância de uma compreensão abrangente, informada e ética dos algoritmos de busca, delineando um percurso significativo para o futuro da CI e sua interseção com as tecnologias digitais.

6 CONSIDERAÇÕES FINAIS

Ao chegarmos ao ponto culminante desta jornada de investigação, é imperativo reservar um momento para sintetizar os *insights*, avanços e implicações que emergiram ao longo deste estudo, que emerge um panorama enriquecedor e reflexivo. A jornada empreendida, guiada pelas diretrizes da DSR revelou não apenas a complexidade dos processos de automação por trás dos motores de busca, mas também a maneira pela qual essa complexidade está intrinsecamente ligada aos princípios e à evolução da CI.

A busca pela compreensão dos Algoritmos de Motores de Busca não se limitou a uma exploração técnica, mas mergulhou nas raízes teóricas que sustentam essa disciplina multidisciplinar. Ao entender os alicerces da CI, pudemos iluminar os meandros de como os algoritmos operam, como selecionam e classificam informações, e como moldam a interação humana com o vasto universo digital.

A pesquisa em estudo, intitulado “Aparato dos Algoritmos na Ciência da Informação: Compreendendo os Processos de Automação dos Motores de Busca”, revelou uma análise profunda e abrangente sobre os algoritmos que operam nos motores de busca, considerando sua influência na recuperação e gestão da informação, bem como os desafios éticos e técnicos associados a esses sistemas automatizados.

O processo de "Entendimento das Necessidades" é fundamental para garantir que um projeto ou sistema atenda aos requisitos, objetivos e expectativas. Dessa forma, no contexto da gestão de informação, a compreensão é essencial para a coleta, organização, armazenamento e disseminação eficazes das informações relevantes.

Este processo envolve interações colaborativas com os *stakeholders* para obter uma visão abrangente das necessidades e requisitos. Essa compreensão profunda é crucial para criar sistemas de informação que proporcionem valor aos usuários e melhorem a tomada de decisões.

Em projetos de tecnologia da informação, como o algoritmo de aplicação *web de web scraping*, o entendimento das necessidades orienta o desenvolvimento de funcionalidades relevantes, a seleção de fontes de dados adequadas e o *design* de interfaces amigáveis. Isso é fundamental para o sucesso do projeto.

No contexto deste estudo, exploramos profundamente como os algoritmos desses motores de busca impactam a recuperação e gestão da informação. Ao considerar a compreensão das necessidades como uma etapa crítica, podemos orientar o desenvolvimento de projetos, como a aplicação *web de web scraping*, em direção ao sucesso, ancorados no cerne

da gestão de informação. Isso não apenas fortalece a implementação de tecnologias, mas também impulsiona uma cultura de gestão informacional eficiente e direcionada.

A pesquisa realizada neste estudo contribuiu significativamente para o campo da CI. Embora não façamos referência a autores específicos, é possível destacar como nossa pesquisa se relaciona com este campo. Exploramos a fundo os algoritmos de busca, investigando seu impacto na gestão de informações, abordando os desafios éticos e técnicos associados. Essa abordagem fornece uma base sólida para discussões esclarecedoras e o desenvolvimento de estratégias futuras no âmbito da CI.

No que diz respeito aos critérios de validade na CI, nossa exploração enriquece a compreensão da importância da transparência algorítmica e da mitigação de vieses, elementos cruciais na construção de sistemas mais justos e imparciais. Tais considerações estão em linha com as reflexões sobre normas e algoritmos na ciência e na informação.

A aplicação do algoritmo de *web scraping* desempenha um papel fundamental, automatizando a coleta de informações de páginas *web*. Essa funcionalidade está diretamente relacionada ao campo da RI, um elemento vital na CI. Nesse campo, o objetivo principal é extrair e organizar conhecimento a partir de diversas fontes. Os resultados são apresentados de forma estruturada e categorizada, destacando a importância de uma organização clara. Além disso, a aplicação oferece aos usuários um acesso conveniente a informações relevantes por meio de uma interface *web*, alinhando-se com a missão da CI de facilitar o acesso e a disseminação eficaz de informações. A capacidade de buscar um termo específico e visualizar os resultados exemplifica os processos de RI, que têm como objetivo encontrar e apresentar informações pertinentes em um conjunto diversificado de recursos.

Ao lançarmos um olhar para o porvir, nos deparamos com perspectivas empolgantes e promissoras que aguardam ser exploradas. O campo da pesquisa em torno dos algoritmos de busca revela horizontes ainda vastos e inexplorados. Os caminhos que se delineiam para futuras investigações vislumbram a expansão da análise empírica, o refinamento de soluções algorítmicas que priorizem a transparência e a inclusão, e a criação de ferramentas educacionais voltadas a capacitar o público geral na compreensão e navegação discernida na era da informação.

Essas direções futuras ecoam as considerações cujo estudo, enfoca a estratégia empresarial e as tecnologias da informação na gestão do conhecimento técnico-documental. Assim como eles examinaram a intersecção entre estratégia e tecnologia, essa pesquisa olha para a interseção entre algoritmos de busca, recuperação da informação e a compreensão de

motores de busca de informação na *web*, ambos os enfoques visando enriquecer a maneira como as organizações e a sociedade como um todo interagem com a informação.

As abordagens inovadoras, aqui proposta, assim como esta pesquisa, os algoritmos de busca compartilham uma essência comum: a busca pela compreensão e otimização dos processos informacionais. No mesmo viés, as direções delineadas para o futuro não apenas ampliarão nosso entendimento dos algoritmos de busca, mas também moldarão o desenvolvimento dessas tecnologias, tornando-as mais transparentes, inclusivas e eficazes.

Na lógica deste artefato, o algoritmo desenvolvido coleta dados de páginas *web* relacionados a um termo de pesquisa, exemplificando uma atividade fundamental na gestão de informação, que abrange a captura e coleta de dados de várias fontes para análise posterior. A utilização de um banco de dados (MySQL) para armazenar informações coletadas destaca a importância da gestão adequada de dados, facilitando a recuperação e administração das informações ao longo do tempo. A aplicação oferece uma interface para que os usuários acessem e utilizem as informações coletadas, alinhando-se com a gestão da informação que tem o propósito de fornecer dados relevantes para aprimorar a tomada de decisões.

A formatação dos resultados e a distinção entre informações relevantes e cabeçalhos exemplificam o processamento e apresentação de dados, sendo a apresentação eficaz e crucial para tornar a informação compreensível e utilizável. Dessa forma, o algoritmo ilustra a aplicação prática da CI e Gestão de Informação ao utilizar técnicas como *web scraping*, organização de dados, recuperação da informação e apresentação de resultados. Assim, exemplifica como os conceitos e métodos desses campos podem ser aplicados para transformar dados brutos em informações úteis para os usuários finais.

Através da incorporação da pesquisa científica e da aplicação prática, uma abordagem que é a essência da DSR, foi impulsionada tanto o entendimento teórico quanto o exploratório, com um empenho incansável em construir conhecimento aplicado e de valor palpável. As conclusões traçadas por esta pesquisa concederam visões amplas e bem embasadas sobre os algoritmos dos motores de busca e a influência que exercem na gestão da informação. Alicerçada sobre os alicerces sólidos da CI, esta investigação dissipou a aura de mistério em torno dos algoritmos de busca, bem como sua aplicação na RI. A análise meticulosa de modelos teóricos, a revisão crítica da literatura e a exploração das implicações éticas culminaram em um conjunto robusto de conhecimentos.

O desiderato desta pesquisa, em consonância com os métodos estabelecidos, revelou-se exitoso na consecução de seus objetivos delineados. Com diversas e minuciosas análises, a pesquisa conseguiu obter um conhecimento profundamente detalhado e perspicaz

sobre o comportamento dos algoritmos nos processos de automação dos motores de busca, especialmente sob a lente perscrutadora da Ciência da Informação (CI).

Em sua essência, o escopo da pesquisa abraçou uma abordagem metodológica robusta, a Design Science Research, que proporcionou um arcabouço teórico-prático imbuído de profundidade e substância. Os resultados obtidos não apenas contribuem para o avanço do campo científico em CI, mas também carregam implicações práticas que reverberam no âmbito profissional e na sociedade contemporânea como um todo.

No que diz respeito aos objetivos específicos, a investigação detalhada das estruturas dos algoritmos utilizados nos motores de busca resultou em uma compreensão refinada da complexidade de sua composição. A análise abrangente dos mecanismos de busca na web revelou os fatores e sinais considerados pelos algoritmos, destacando variáveis que influenciam a relevância e a ordenação dos resultados de pesquisa. Essa exploração minuciosa proporcionou uma visão detalhada das nuances envolvidas nos processos decisórios dos algoritmos.

A concretização do objetivo de desenvolver e implementar um exemplo de Web Crawler, utilizando a linguagem de programação Python, foi crucial para a compreensão da interação entre o processo de automação e os algoritmos de busca. A realização de buscas automatizadas de informações na web, por meio desse Web Crawler, ofereceu uma experiência prática que enriqueceu a compreensão teórica, proporcionando uma visão holística da dinâmica envolvida.

Ao considerar os resultados obtidos em cada objetivo, é evidente que esta pesquisa não apenas cumpriu suas metas, mas também transcendeu expectativas, fornecendo contribuições substanciais para o campo científico. A interdisciplinaridade entre teoria e prática, aliada à abordagem metodológica, estabeleceu uma base para futuras explorações e inovações na compreensão dos algoritmos de busca na CI. Essa pesquisa não é apenas um marco no entendimento dos motores de busca, mas também representa um convite para investigações mais aprofundadas e reflexões contínuas sobre o papel desses algoritmos na sociedade contemporânea.

A meticulosa execução dos objetivos específicos proporcionou uma análise profunda dos algoritmos de busca, contribuindo para o desenvolvimento de um arcabouço teórico-prático. A compreensão reveladora da complexidade estrutural e do funcionamento intrincado desses algoritmos amplia não apenas o conhecimento acadêmico na área de CI, mas também oferece uma compreensão para a aplicação prática em diversos setores.

Ao desvelar os fatores e sinais que influenciam as decisões algorítmicas na classificação e apresentação dos resultados de pesquisa, a pesquisa preenche uma lacuna crucial na compreensão desses sistemas complexos. A análise abrangente dos mecanismos de busca na web fornece uma visão profunda das variáveis envolvidas, permitindo uma compreensão mais tangível e aplicada dessas tecnologias essenciais na sociedade contemporânea.

A implementação bem-sucedida do exemplo de Web Crawler, desenvolvido em Python, não apenas ilustra a aplicação prática dos conhecimentos adquiridos, mas também destaca a importância da interação entre o processo de automação e os algoritmos de busca. A experiência prática adquirida por meio desse processo fortaleceu não apenas a compreensão teórica, mas também proporcionou uma perspectiva valiosa sobre os desafios e possibilidades na implementação dessas tecnologias.

Nesse contexto, é imperativo ressaltar que tais realizações não são somente intrínsecas ao âmbito da CI, mas também convergem com as demandas prementes da atualidade. A sociedade contemporânea, imersa na era da informação, confronta-se com desafios e oportunidades proporcionados pela crescente automatização dos processos de busca de informação. Nesse cenário, a compreensão profunda dos algoritmos de busca assume um papel preponderante, uma vez que esses algoritmos atuam como os árbitros que determinam como a informação é selecionada, classificada e disponibilizada aos ávidos buscadores de conhecimento.

Consequentemente, a pesquisa não somente propicia uma compreensão aprofundada dos processos algorítmicos nos motores de busca, mas também realça a importância da transparência algorítmica e da mitigação de vieses. Em um contexto no qual a justiça e a imparcialidade na disponibilização da informação são inegociáveis, os *insights* derivados desta pesquisa não apenas enriquecem o campo da CI, mas também apresentam considerações fundamentais para a sociedade que demanda uma interação ética e equitativa com a informação.

A aplicação prática do algoritmo desenvolvido, permitindo a coleta de dados de páginas web, exemplifica de maneira vívida a importância da gestão adequada de dados, o que culmina na facilitação da recuperação e administração das informações ao longo do tempo. A apresentação estruturada e categorizada dos resultados, diferenciando informações pertinentes de cabeçalhos, ilustra de forma eloquente o processamento e a apresentação de dados, que são elementos vitais para tornar a informação compreensível e utilizável. Assim, o algoritmo concebido não apenas corrobora os princípios da CI e da Gestão da Informação, mas também

comprova como esses conceitos e métodos podem ser aplicados para transformar dados brutos em informações de valor inestimável para os usuários finais.

Este estudo não apenas atingiu seus objetivos, mas também identificou áreas promissoras para investigações futuras. A análise contínua dos mecanismos de busca, visando aprimorar a recuperação e o gerenciamento de informações, e a exploração para redução de viés nos algoritmos de busca, permanecem como campos ricos para aprofundamento. O crescente reconhecimento dos impactos sociais e éticos dessas tecnologias destaca a necessidade constante de adotar abordagens críticas e responsáveis.

Para além de ser uma conquista significativa no âmbito da CI, esta pesquisa fomenta uma compreensão mais profunda dos algoritmos de busca. O caminho delineado por este estudo estabelece uma base para o progresso contínuo na interseção entre algoritmos e Ciência da Informação, destacando a importância de uma abordagem abrangente, bem informada e ética para moldar o futuro dessas tecnologias.

À guisa de uma última reflexão, este estudo destaca o horizonte promissor e empolgante que se descortina no campo da pesquisa em algoritmos de busca. A busca pela compreensão e otimização dos processos informacionais, enraizada na Ciência da Informação, não apenas ampliou nosso entendimento dos algoritmos de busca, mas também pavimentou o caminho para o desenvolvimento de tecnologias mais transparentes, inclusivas e eficazes. As direções futuras delineadas são portadoras de uma promissora continuidade na pesquisa, que poderá desvelar novos panoramas e possibilidades no âmbito da Ciência da Informação.

REFERÊNCIAS

- ALBAGLI, S. Ciência aberta em questão. *In*: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciências Abertas, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO. 2015. p. 9-27.
- ALMEIDA, M. F.; BASTOS, F. C.; BITTENCOURT, G. A produção do conhecimento em Ciência da Informação no Brasil. *In*: Encontro Nacional de Pesquisa em Ciência da Informação, 8., 2007, Salvador. **Anais [...]**, Salvador: ANCIB, p. 69-77. 2007.
- ARAÚJO, C. A. Á. **O que é ciência da informação**. Belo Horizonte: KMA, 2018.
- BARANAUSKAS, M. C. C. *et al.* Design Science Research: método de pesquisa para a engenharia de produção. **Produção**, v. 19, n. 3, p. 450-462, p. 2009.
- BARBOSA, R. R. **Modelagem de Processos de Negócio**: Uma Abordagem Prática. Rio de Janeiro: Novatec Editora, 2018.
- BARRETO, A. M. O uso do computador na Ciência da Informação no Brasil: um estudo bibliométrico. **Perspectivas em Ciência da Informação**, [S.l.], v. 12, n. 2, p. 56-67, 2007.
- BATISTA, M. **Gestão da Informação**: Como transformar informações em ativos para a empresa. São Paulo, Editora Atlas, 2014.
- BAWDEN, D.; ROBINSON, L. **Introduction to information science**. [S.l.]: Facet Publishing, 2012.
- BAX, M. P. Design Science Research Contributions: Mapping the Structure of the Field. **Journal of the Association for Information Systems**, v. 16, n. 12, p. 1058-1087, 2015.
- BECKER, H. Observação social e estudos de casos sociais. *In*: **Métodos de pesquisa em Ciências Sociais**. São Paulo: Hucitec, 1999.
- BORSATO, M. C. **Metodologia científica**. Rio de Janeiro: Pearson Brasil, 2013.
- BRAGA, L. M. **Ciência da Informação e Biblioteconomia**: Distinções e Relações. Niterói: EdUFF, 1995.
- CAPURRO, R.; HJORLAND, B. The concept of information. **Annual Review of Information Science and Technology**, v. 41, n. 1, p. 343-411, 2007.
- CERAVOLO, S. M.; TÁLAMO, M. F. G.M. Tratamento e organização de informações documentárias em museus. **Revista do Museu de Arqueologia e Etnologia**, São Paulo, v. 10, p. 241-253, 2000.
- CORMEN, T. H. **Algoritmos - Teoria e Prática**. Rio de Janeiro, GEN LTC, 2010.
- CORMEN, T. H. *et al.* **Introduction to Algorithms**. 3. ed. Massachusetts: MIT Press, 2009.

CUNHA, M. B. da. **Metodologias ativas de aprendizagem**: guia prático para professores de ensino superior. Editora Vozes, 2017.

DECARLI, G.; TSAGARAKI, C. **Un Inventário de Bienes Culturales**: ¿por qué y para quién?. ILAM – Instituto Latino Americano de Museus. San José, Costa Rica: Ediciones ILAM, 2006.

DIAS, C. M. Paradigmas em ciência da informação. **DataGramaZero**, Rio de Janeiro, v. 3, n. 3, 2002.

DIAS, P. A evolução dos estudos de Ciência da Informação. **Ciência da Informação**, Rio de Janeiro, v. 31, n. 2, p. 7-18, 2002.

DRESCH, A.; LACERDA, D. P.; ANTUNES JR, J. A. V. **Design Science**: Método de Pesquisa em Engenharia de Produção e Gestão de Operações. Atlas. São Paulo, 2015.

EGGHE, L.; ROUSSEAU, R. Introduction to informetrics: quantitative methods in library, documentation and information science. In: EGGHE, L.; ROUSSEAU, R. **Introduction to informetrics**: quantitative methods in library, documentation and information science. Amsterdam: Elsevier, 1990, p. 15-30.

FERREZ, H. D. Documentação museológica: teoria para uma boa prática. In: Cadernos de ensaios, n. 2. **Estudos de museologia**. Rio de Janeiro: Minc/Iphan, p. 64- 67, 1994.

FOLTA, J. Some remarks on the history of numerical analysis especially in the area of Prague. **Llull**, v. 11, p. 217–233, 1988.

FREIRE, I. M. Da construção do conhecimento científico à responsabilidade social da Ciência da Informação. **Informação & Sociedade**: estudos. João Pessoa, v. 12, n. 1, 2002. Disponível em: <https://ridi.ibict.br/bitstream/123456789/194/1/FREIREIS2002.pdf>. Acesso em: 3 de jan. 2023.

GIBBONS, M. *et al.* The new production of knowledge: The dynamics of science and research in contemporary societies. **SAGE**, 1994. Disponível em: <https://sk.sagepub.com/books/the-new-production-of-knowledge>. Acesso em: 1 out. 2023.

GLUCHOFF, A. Artillerymen and mathematicians: Forest Ray Moulton and changes in American exterior ballistics, 1885–1934. **Hist. Math.**, v. 38, n. 4, p. 506–547, 2011.

GOFFMAN, E. **Encounters**: Two Studies in the Sociology of Interaction. Rio de Janeiro: Bobbs-Merrill, 1970.

GOLDSTINE, H. H. **A History of Numerical Analysis**. From the 16th Through the 19th Century. New York: Heidelberg, 1977.

GONÇALVES, M. A.; FOX, E. A. Digital libraries and web search engines: The algorithms under the hood. **Journal of the Association for Information Science and Technology**, v. 70, n. 3, p. 239-243, 2019.

GOMEZ, G. M. N. O campo científico da Ciência da Informação. **Informação & Informação**, v. 6, n. 1, p. 51-66, 2001.

GOOGLE SEARCH CENTRAL. **In-depth guide to how Google Search works**. 2022. Disponível em: <https://developers.google.com/search/docs/fundamentals/how-search-works>. Acesso em: 3 jan, 2022.

GUIMARÃES, J. A. C. **Análise SWOT**: desenvolvimento de ferramenta para tomada de decisões em gestão de portfólio de projetos. 2020. Dissertação (Mestrado) - Universidade de São Paulo, São Paulo, 2020.

GREGOR, S.; HEVNER, A. R. Positioning and presenting design science research for maximum impact. **MIS quarterly**, v. 37, n. 2, p. 337-355, 2013.

Harari, Yuval Noah. **21 Lições para o Século 21**. Companhia das Letras, 2019.

HEVNER, A.; CHATTERJEE, S. **Design research in information systems**: theory and practice. [S.l.]: Springer Science & Business Media, 2010.

HEVNER, A. R. *et al.* Design science in information systems research. **MIS Quarterly**, v. 28, n. 1, p. 75-105. 2004.

JANSEN, B. J.; BOOTH, D. L.; SPINK, A. Determining the informational, navigational, and transactional intent of Web queries. **Information Processing & Management**, v. 44, n. 3, p. 1,251–1,266, 2008.

KOFFMAN, E.B.; FRIEDMAN, F. L. **FORTTRAN**: fifth edition update. Massachusetts: Inc. Addison Wesley Longman, 1997.

KNUTH, D. E. **The art of computer programming**, v. 1: Fundamental. Algorithms. Reading, MA: Addison-Wesley, 1968.

KUHLMAN, D. **A Python Book**: Beginning Python, Advanced Python, and Python Exercises. 2012. Disponível em: https://web.archive.org/web/20120623165941/http://cutter.rexx.com/~dkuhlman/python_book_01.html#id25 Acesso em: 10 jan. 2023.

LARA, L. L. de *et al.* **Modelagem de Sistemas de Informação**. São Paulo: Editora Novas Edições Acadêmicas, 2019.

LE COADIC, Y. F. **A Ciência da Informação**. Brasília: Briquet de Lemos/Livros, 2004.

LEISERSON, C. E. *et al.* **Algoritmos**: teoria e prática. Rio de Janeiro: Campus, 2009.

LESK, M. **Understanding Digital Libraries**. The Morgan Kaufmann Series in Multimedia Information and Systems. [S.l.]: Elsevier, 2004.

LEVENE, M. An Introduction to search Engines and web navigation. **John Wiley & Sons**, Inc, Publication. 2010. Disponível em: <https://srikarthiks.files.wordpress.com/2016/07/t4-an-introduction-to-search-engines-and-web-navigation-2nd.pdf>. Acesso em: 3 jan. 2023.

LÉVY, P. **As tecnologias da inteligência**: o futuro do pensamento na era da informática. Rio de Janeiro: Editora 34, 1993.

LIU, X.; ZHU, J. The Influence of Recommendation Algorithms on Search Engine Results and User Experience: a literature review. **Information Research**, v. 26, n. 1, 2021.

LOUREIRO, A. A. F. «**Análise de complexidade**» Pós-graduação em ciência da computação. Slides. 2010. Disponível em: <http://www.decom.ufop.br/menotti/paa101/slides/aula-AnaliseAlgoritmos.pdf>. Acesso: 15 jun. 2023.

MARCHAND, D. A.; DAVENPORT, T. H. **Mastering Information Management**. New Jersey: Financial Times Prentice Hall, 2000.

MATTELART, A. **História da sociedade da informação**. São Paulo: Loyol, 2005.

MEADOWS, J. **A comunicação científica**. Brasília: Briquet de Lemos/Livros, 1999.

MENEZES, N. N. C. **Introdução à Programação com Python** - Algoritmos e lógica de programação para iniciantes. 2. ed. São Paulo, Nova tec Editora Ltda, 2014.

MENEZES, P. B. **Linguagens Formais e Autômatos**. 4. ed. Porto Alegre: Sagra-Luzzato, 2001. (Livros Didáticos, v. 3).

MIKHAILOV, A. I.; GILJAREVSKIJ, R. S. On Information Science. **Journal of Documentation**, v. 26, n. 1, p. 14, 1970.

MIRANDA, A. Ciência da Informação: história, conceito e características. **Em Questão**, v. 21, n. 3, p. 25-42, 2002.

MIRANDA, A. R. A ciência da informação como disciplina acadêmica. **Perspectivas em Ciência da Informação**, v. 7, n. 1, p. 7-25, 2002.

MORAES, E. A. A prática da pesquisa em ciência da informação. In: MORAES, E. A.; MENEGHINI, R. (Eds.). **Metodologias ativas para uma educação inovadora**: uma abordagem teórico-prática. São Paulo: Editora Senac, 2018.

MUELLER, J. P.; MASSARON, L. **Algoritmos para leigos**. Traduzido por Ana Araujo. Rio de Janeiro: Alta Books, 2018.

MUELLER, J.; MASSARON, L. **Algorithms For Dummies**. New Jersey: John Wiley & Sons, 2017.

OLIVEIRA, M. G. V. **Biblioteca pública**: do controle ao acesso. Brasília: Briquet de Lemos, 2005.

O'Neil, C. **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. Broadway Books, 2016.

ORTEGA, C. D.; LARA, M. L. G. de. A noção de documento: de Otlet aos dias de hoje. **DataGramaZero**: revista de Ciência da Informação, v. 11, n. 2, 2010.

PACKER, A. N. **Compreendendo sistemas de informação**. Rio de Janeiro: Bookman Editora, 2010.

PEFFERS, K. *et al.* (2020). A design science research methodology for information systems research. **Journal of management information systems**, v. 37, n. 1, p. 9-45, 2020.

PEJOVIC, D. Application of Python Programming Language in Measurements. **Electronics and Energetics**, v. 32, n. 1, p. 1-23, mar. 2019. Disponível em: <http://www.doiserbia.nb.rs/img/doi/0353-3670/2019/0353-36701901001P.pdf>. Acesso em: 12 jan. 2023.

PINHEIRO, L. V. R. A contribuição da bibliometria para a Ciência da Informação. **Informação & Sociedade**, João Pessoa, v. 7, n. 2, p. 63-79, 1997.

PINHEIRO, L. V. R. **A Ciência da informação entre sombra e luz**: domínio epistemológico e campo interdisciplinar. 2002. Tese (Doutorado) - Universidade Federal de Minas Gerais, Minas Gerais, 2002.

PINHEIRO, L. V. R. A documentação e a informação científica: evolução, definição e campo de atuação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 7, n. 2, p. 159-174, 2002.

PINHEIRO, L. V. R. O objeto de estudo da Ciência da Informação. In: LANDIM, R. C. G.; PINHO, M. C. R. de; OLIVEIRA, M. G. V. (Orgs.). **Introdução à Ciência da Informação**. Brasília: Thesaurus, 2002.

PINHEIRO, L. V. R. Processo evolutivo e tendências contemporâneas da Ciência da Informação. **Informação & Sociedade: Estudos**, v. 15, n. 1, p. 13-48, 2005. Disponível em: <https://brapci.inf.br/index.php/article/view/0000007733/e672decab77f7bd84b92e05e5>. Acesso em: 23 de nov. 2022.

PRIES-HEJE, J.; BASKERVILLE, R.; VENABLE, J. Strategies for design science research evaluation. In: **Proceedings of the 16th Australasian Conference on Information Systems** (ACIS 2005). Sydney, Australia: [s.n.], 2005.

RAMALHO, L. **Fluent Python**. 2. ed. [S.l.]: O'Reilly Media, 2022.

RAYWARD, W. B. The origins of information science and the International Institute of Bibliography/International Federation for Information and Documentation (FID). **Journal of the American Society for Information Science**, 1997. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(199704\)48:4%3C289::AID-ASI2%3E3.0.CO;2-S](https://doi.org/10.1002/(SICI)1097-4571(199704)48:4%3C289::AID-ASI2%3E3.0.CO;2-S). Acesso em: 28 dez. 2022.

RIBEIRO, A. C. M. L.; SANTOS C. D. Isso não é uma pirâmide: revisando o modelo clássico de dado, informação, conhecimento e sabedoria. **Revista Ciência Da Informação**, Brasília: IBICT, v. 49, n. 2, p. 67- 87, 2020.. Disponível em: <http://revista.ibict.br/ciinf/article/view/5066>. Acesso em: 28 dez. 2022.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. Why Should I Trust You?: Explaining the Predictions of Any Classifier. **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, 2020. Disponível em: <https://doi.org/10.1145/2939672.2939778>. Acesso: 1 out. 2023.

RIEMENSCHNEIDER, C. K.; HARDGRAVE JR, B. C.; DAVIS, F. D. **Explaining software development tool use**: An extended task-technology fit model. [S.l.]: MIS Quarterly, 106-127, 2002.

RIEMENSCHNEIDER, C. K.; HARDGRAVE, B. C.; DAVIS, F. D. Explaining Software Developer Acceptance of Methodologies: A Comparison of Five Theoretical Models. **IEEE Transactions on Software Engineering**, Nashville, TN, USA, v. 28, n. 12, p.1135-1145, 2002.

RUSSO, M. A. **Informação e conhecimento científico**: estudo de suas relações na contemporaneidade. 2010. Dissertação (mestrado) - Universidade Federal de Santa Catarina, Santa Catarina, 2010.

RUSSO, P. C. Os periódicos científicos e o desenvolvimento da ciência: uma breve reflexão histórica. **Ponto de Acesso**, São Carlos, v. 4, n. 1, p. 63-76, 2010.

SARACEVIC, T. Ciência da Informação: origem, evolução e relação com a biblioteconomia. **Perspectivas em Ciência da Informação**, Rio de Janeiro, v. 1, n. 1, p. 41-62. 1996.

SEDGEWICK, R.; WAYNE, K. **Algorithms**. 4. ed. Boston, MA, USA: Addison-Wesley, 2011.

SILVA, C. L. da. **Ciência da Informação**: conceitos e fundamentos. Rio de Janeiro: E-papers, 2006.

SILVA, J. L. C.; FREIRE, G. H. de A. Um olhar sobre a origem da Ciência da Informação: indícios embrionários para sua caracterização identitária. Florianópolis, **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, 2012. Disponível em: <http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n33p1/21708> , Acesso em: 29 dez. 2022.

SIMON, H. A. **The sciences of the artificial**. Cambridge, MA, USA: MIT Press, 1996.

SINGH, K.; KANT, R. Personalized search and its impact on search engine results quality: A review. **Information Processing & Management**, Amsterdam, Netherlands, v. 57, n. 4, p. 102-207, 2020.

SOLANA-GONZÁLEZ, P.; PÉREZ-GONZÁLES, D. Estrategia empresarial y tecnologías de la información en la industria moderna: Un enfoque de estudio de caso en la empresa XYZ. **Revista de Gestión Tecnológica**, v. 20, n. 2, p. 45-62, 2021.

SOUZA, C. A. **Design Science Research em Sistemas de Informação e Engenharia de Software**: Conceitos e Práticas. São Paulo: Novatec Editora, 2019.

SPITZNAGEL, E. L. **Selected topics in mathematics**. New York: Holt, Rinehart and Winston, 1971.

TAVARES, P. de C. **Algoritmo**, in "Enciclopédia Verbo Luso-Brasileira da Cultura, Edição Século XXI", Volume II. Rio de Janeiro: Editorial Verbo, Braga, 1998.

TAYLOR, A. G. **A glossary of library terms**. London: André Deutsch, 1951.

TOCCI, R. J.; WIDMER, N. S.; MOSS, G. L. **Sistemas Digitais: Princípios e Aplicações**. 10. ed. São Paulo: Pearson, 2007.

TOSCANI, L.; VELOSO, P. **Complexidade de Algoritmos: análise, projetos e metodos**. Porto Alegre: SAGRA-Luzzatto, 2001. (Livros Didáticos, v. 13).

TOURNES, D. **La construction tractionnelle des équations différentielles**. Paris: Blanchard, 2009.

TSATSOU, P. *et al.* A SWOT analysis of web-based library services. **Library Management**, v. 32, n. 6/7, p. 457-471, 2011.

TUOMINEN, K.; SAVOLAINEN, R.; TALJA, S. **Information Science as a Multidisciplinary Field**. In The Oxford Handbook of Information Science [S. l.]: Oxford University Press, 2018.

XIE, H.; LI, X. Search Engine Bias and the Demise of SearchEngine Utopianism. **Social Media + Society**, v. 4, n. 3, 2018.

YIN, R. **Case Study Research: design and methods**. 5. ed. Thousand Oaks, CA: Sage, 2014.

ZHANG, Y.; XU, L.; SUN, H. **Baidu: A Benchmark for Evaluating the Effectiveness of Learning to Rank Methods in Chinese**. [S.l.]: Proceedings of the 42nd International ACM, 2019

ZIVIANI, N. **Projeto de algoritmos: com implementações em Pascal e C**. Rio de Janeiro: Cengage Learning, 2011.